PREDICTING DIGITAL CURRENCY MARKET WITH SOCIAL DATA: IMPLICATIONS OF NETWORK STRUCTURE AND INCENTIVE HIERARCHY

A Dissertation
Presented to
The Academic Faculty

by

Peng Xie

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Scheller College of Business

Georgia Institute of Technology
August 2017

PREDICTING DIGITAL CURRENCY MARKET WITH SOCIAL DATA: IMPLICATIONS OF NETWORK STRUCTURE AND INCENTIVE HIERARCHY

Approved by:

Dr. Yu Jeffrey Hu
Scheller College of Business
Georgia Institute of Technology

Dr. Hailiang Chen
College of Business
City University of Hong Kong

Dr. Sabyasachi Mitra
Scheller College of Business
Georgia Institute of Technology

Dr. Eric Overby
Scheller College of Business
Georgia Institute of Technology

Dr. Lizhen Xu
Scheller College of Business
Georgia Institute of Technology

Date Approved: July 18, 2017

ACKNOWLEDGEMENTS

I am very grateful to my family. They provided everything necessary in my pursuit of the degree. They always try to learn my progress and help me through all the difficulties. They flew all the way across the ocean when I was at the all-time low. I could not have reached this stage without their supports along the way.

I would like to express my sincere appreciation to my major advisor Dr. Jeffrey Hu. He is so keen on finding interesting research opportunities, spotting problems in my research, and suggesting methods to improve. After working with him for almost three years, I definitely became a more rigorous person in every way.

I am also extremely thankful to Dr. Hailiang Chen. He taught me the right way to approach a new research question and walked me through each step. He is extremely smart and helpful. I still remember how I was impressed when he sent me the revised manuscripts with so many notes and marks that cover every tiny detail of the paper.

I also gratefully acknowledge the support from the committee members: Dr. Sabyasachi Mitra, Dr. Eric Overby, Dr. Lizhen Xu, all the ITM department faculty members: Dr. Sridhar Narasimhan, Dr. Marius Florin Niculescu, Dr. Michael Smith, Dr. D. J. Wu, Dr. Han Zhang, and the Scheller staff Ursula Reynolds and Shannon Smith. Thank you so much for your supports and efforts to make my study here possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

SYMBOLS

P_t	Time t Bitcoin price
R_t	Time t Bitcoin return or Altcoin return
$Sentiment_{it}$	The sentiment for Altcoin *i* at time *t*
X	Control variables
η_{it}	Linear regression error term
α_t	Linear regression time fixed effect
$\log(TradingVolume)_t$	Logarithm of Bitcoin trading volume at time *t*
$\log(PostCount)_{it}$	Logarithm of time *t* Bitcoin discussion post count in thread *i*
ε_{it}	The regular disturbance in the linear regression error term
u_i	The thread specific disturbance in the linear regression error term
σ_ε^2	Variance of regular disturbance
$\sigma_{u_i}^2$	Variance of thread specific disturbance
$Cohesion_{it}$	The connectedness measure for the discussion network in thread *i* during day *t*
Sentiment-Broadcast$_{it}$	The sentiment of all broadcast messages for the discussion network in thread *i* during day *t*
Sentiment-Discussion$_{it}$	The sentiment of all discussion messages for the discussion network in thread *i* during day *t*
IndegreeCentrality$_{it}$	The in-degree centrality for author *i* during day *t*
OutdegreeCentrality$_{it}$	The out-degree centrality for author *i* during day *t*
OutdegreeCentrality$_{it}$	The Bitcoin trading intensity during day *t*
Averagedegree$_{it}$	The average degree for the discussion network in thread *i* during

	day *t*
$Density_{it}$	The density for the discussion network in thread *i* during day *t*
$HRSentiment_{it}$	Aggregated sentiment of high rank social media users about Altcoin *i* during day *t*
$LRSentiment_{it}$	Aggregated sentiment of low rank social media users about Altcoin *i* during day *t*
$BtcHRSentiment_{t}$	Aggregated sentiment of low rank social media users about Bitcoin *i* during day *t*
$BtcLRSentiment_{it}$	Aggregated sentiment of low rank social media users about Bitcoin *i* during day *t*
$BTCR_t$	Bitcoin return during day *t*
$ALTR_t$	Altcoin *i*'s return during day *t*

ABBREVIATIONS

# Obs.	The number of observation
Altcoin	All non-Bitcoin cryptocurrency

SUMMARY

Following the recent discovery of social media's predictive power for the financial markets, we try to advance the literature by finding ways to distinguish between value-relevant information and noises by investigating the implication of social media network structures and its incentive hierarchy system. In the first chapter, we use data from the Bitcoin market and empirically show that highly-connected social media discussion networks are less accurate in predicting future returns due to information free riding and highly correlated information. However, social media information linkages nevertheless serve as landmarks for identifying informed social media actors: value-relevant information is more likely to be shared by users who stimulate active discussions among their peers. In the second chapter, we examine how social media incentive hierarchy systems shape users' posting motivation and thus influence the information quality. We find that on average, the active social media users holding higher badges provide less accurate prediction compared to inactive users with lower badges. The reduced motivations after obtaining the higher level badges and the more frequent use of social media for socialization purposes imply a higher proportion of noises in their posts.

NETWORK STRUCTURE AND PREDICTIVE POWER OF SOCIAL MEDIA IN BITCOIN MARKET

Financial market investors traditionally rely on information sources such as company disclosures, market news, analyst reports and so on to make investment decisions. However, as social media rapidly penetrates into various aspects of our life over the past few years, it also demonstrates the possibility to serve as an alternative information outlet for financial market investors. Rather than resorting to traditional information providers, investors now have the option to directly communicate with peers and exchange investment ideas through social media.

Related academic works have already empirically shown social media's potential in providing financial market participants with value-relevant information and in aiding them to form investment views (e.g., Antweiler and Frank 2004; Chen et al. 2014; Das and Chen 2007; Tumarkin and Whitelaw 2001). Industrial practitioners have also realized the value of social media in recent years and experimented new methods to conduct financial analysis by monitoring the dynamics of social media. Examples include DataSift (a social data company that collects and analyzes unstructured data from social media sites such as Facebook), and Cayman Atlantic (an investment management company that operates Managed Trading Accounts based on real-time event detection with data from social media such as Twitter, Facebook and other similar channels using sentiment analysis).

However, extracting value-relevant information from social media platforms is not a trivial task. “Social media indicators are valuable, but they are not easy pickings”, as Joe Gits, the founder of Social Media Analytics (SMA) notes that “90% of all the Twitter feeds that SMA analysts dissect are discarded - it’s the other 10% reveal investment opportunities investors are clamoring for”.

This statement points to the rarity of valuable information on social media and the importance of information filtering when resorting to social media for investment advice. These considerations motivate our research. The primary goal of this paper is to distinguish between value-relevant information and noises on social media platforms by examining the relationship between social media network structures and its predictive power. Though network structure is an important property of social media, to our knowledge, researchers have not empirically investigated its role in affecting the predictive power of social media information in financial contexts. We contribute to the literature by filling this gap.

Previous literature has provided theoratical guidelines to our investigations. Han and Yang (2013) modeled the stratigic interactions between rational traders and noise traders and demonstrated that when information acquisition is endogenous (i.e., traders decide on their own whether and when to acquire information), social communication causes information free riding and hurts information production. The result is a decreased fraction of informed traders in the network in equilibrium (Han and Yang 2013). Colla and Mele (2010) modeled information correlations based on a ring lattice information structure and proposed a mechanism that information linkages damage traders’ monoplistic power and traders’ profits as well.

Building on the theoretical foundations, we expect social media predictive power to be negatively associated with its network connectedness due to the existence of information correlation and free riding as the result of information linkages. Information linkages are the local connections that grant socially connected nodes access to the same information. In our research, they are identified based on the quoting behavior on social media sites. When information linkages exist between two messages, they are likely to be focused on the same issue and thus correlated with each other. Such correlated information brings less new information to the discussion network and dilutes the degree of network informativeness.

Information free riding is also responsible for the negative relationship between connectedness and informativeness. When one node quotes another in the network, the quoting node acquires new information from the quoted node. The expectation of learning new information on social media from peers with no search cost gives rise to the information free riding, which hampers the incentive to search and share private information. High network connectedness implies greater potentials for information free riding and therefore a lower proportion of informed actors in the network.

Another stream of research discussing the “hidden profile” effect also offers theoretical supports for our hypothesis. The “hidden profile” effect refers to the tendency to conform to the public information and consensuses held by most group members even if an individual has insightful private information (Qiu et al. 2016; Stasser and Titus 1985). In our case, high density of information linkages implies the existence of common focuses and highly correlated discussions. If the common focuses and the public consensus are overwhelmingly prevailing, fragmentary private information held by each

individual is likely to remain silent and undiscovered during the social interactions. In this scenario, social media discussions fail to piece together each individual's private information and result in an incomplete information set and potentially biased investment opinions. Higher information network connectedness implies a stronger "hidden profile" effect and worse overall network informativeness.

Though information linkages damage the predictive power of information networks, they serve as landmarks for value-relevant information. Receiving many incoming information connections (attracting many discussions from peers) indicates importance and relevance. We hypothesize that valuable investment views are more likely to come from social media participants who trigger a buzz among peers.

In addition to influencing the predictive power for future price movement, network connectedness is also expected to predict future trading intensity. A sudden surge in the volumes of social media communications implies that investors are facing new information on the market. Being uncertain about the impact, they exchange ideas on social media platforms to evaluate the situation and to reduce market risks. The monopoly power of new information diminishes as it spreads across the network through communications, causing a quick wear off of new information and increased trading volumes. Therefore, we propose a positive relationship between the network connectedness of social media discussions and future trading intensity.

Rather than the traditional stock market, we conduct analyses using data from the Bitcoin market. Bitcoin is one of the emerging digital currencies known as cryptocurrency. The advantage of using the Bitcoin market as the research context is the

elevated importance of social media as an information outlet compared to stock markets because there are no official information sources such as company announcements, periodical financial disclosures, and professional financial analysts' opinions in the Bitoin market.

We downloaded social media discussions from a leading message board dedicated to Bitcoin-related topics, Bitcointalk.org, in the period of December 2012 to June 2016. Our sample consists of more than 330,000 messages written by 14,052 unique authors posted on 10,663 different threads. This message board offers a quoting feature allowing users to communicate with each other when posting. Our dataset also captures this communication dynamic.

We first validate social media's predictive power for the Bitcoin market by examining the response of next-day Bitcoin returns to the current-day sentiment extracted from social media discussions, then we incorporate discussion network structures into our models to evaluate its impact on the Bitcoin price response. In the analysis at the network level, we adopt two measures for network connectedness, average degree and density, and investigate whether highly connected networks underperform in predicting the next-day Bitcoin return. In the analysis at the nodal level, we assess whether messages with no quoting are more informative for future price movements than messages creating information linkages. We next examine if individuals attracting many discussions from others provide better predictions. For this purpose, we compute the in-degree centrality and out-degree centrality for each participating author, and evaluate how they affect an author's prediction accuracy. Finally, we regress the next-day trading intensity on the current-day network connectedness to test our last hypothesis.

To preview our results, we find that when the percentage of negative words in discussion networks with an average degree of 0.551 (sample mean) is 1% higher, the next-day Bitcoin return is 0.438% lower. In contrast, the same 1% increase in the percentage of negative words for discussion networks with an average degree of 0.730 (one standard deviation higher from the sample mean) is associated with a decrease of only 0.072% in the next-day Bitcoin return. When network connectedness is measured by density, we obtain similar results that loosely connected networks are more accurate in predicting future returns. A 1% increase in the percentage of negative words in discussion networks with a density of 0.024 (sample mean) is associated with a decrease of 0.504% in the next-day Bitcoin return. For discussion networks with a density of 0.036 (one standard deviation higher from the sample mean), the decrease is only 0.097%.

The author level analysis reveals an interesting pattern that the implication of social media for future price movements comes almost entirely from those who are frequently quoted by their peers. Our empirical evidence shows that when the percentage of negative words is 1% higher in an author's posts, the next-day Bitcoin return is 0.027% lower for an author receiving 1.009 daily quotations (sample mean), whereas the next-day Bitcoin return is 0.061% lower for an author receiving 2.864 daily quotations (one standard deviation higher from the sample mean). Our data suggests that the number of quotations an author gets is highly skewed. Most valuable information comes from the 5% - 10% authors with double-digit number of daily quotations.

Lastly, we find that social media network connectedness predicts future trading intensity. On average, the number of transactions and the trading volume in dollar terms increase by 0.197% and 0.532%, respectively, when network average degree is one

standard deviation higher. Models using network density as the alternative network connectedness measure provide consistent results.

This research mainly contributes to literature of social media's role in financial markets in the following ways: (1) reexamines social media's predictive power for price movements in the context of digital currency; (2) reviews the theoretical foundations for the relationship between network connectedness and its informativeness and empirically tests it; (3) suggests methods to distinguish between informed and uninformed social media actors; and (4) examines the link between social media network structure and future trading intensity.

Apart from theoretial contributions, our research also provides practical implications. We demonstrate that social media network structure can be used to sort out the valuable investment advice within an enormous amount of social data generated each day. Because there is no guarantee for the quality of the information shared on social media, our insights will help reduce information acquisition cost and improve the quality of the information set.

1.1 Literature Review and Hypothesis Development

It has been well documented in the literature that traditional financial reports and editorial media outlets can affect future market price movements. Among these studies, many information channels have been examined. Examples include but are not limited to earnings press release (Davis et al. 2012), 10-Ks (Loughran and McDonald 2011), Wall Street Journal columns (Tetlock 2007), Dow Jones News Service firm-specific news stories (Tetlock et al. 2008), and IR firm news spinning (Solomon 2012). With the rapid

development of social media, researchers started related studies using sentiments extracted from social data in recent years.

One of the biggest differences between traditional financial advice sources and social media is that social media contents are usually loosely organized and informal. There is no guarantee for the quality of information shared on social media platforms, and some researchers did not find strong empirical supports for social media's predictive power. Dewally (2003) used buy and sell recommendations from an online discussion group to predict stock market but failed to establish the relationship. Antweiler and Frank (2004) studied the effects of messages posted on Yahoo! Finance and found only mild influence. However, some other online communities have been shown to successfully predict market movements. Tumarkin and Whitelaw (2001) measured investor opinions on RagingBull.com and found that the investor opinions predict the next-day abnormal returns. Das and Chen (2007) examined Yahoo's message board and documented a relationship with 24 tech-sector stocks. Chen et al. (2014) found that Seeking Alpha articles provide value-relevant information for long-term stock returns.

The existing literature suggests potential for social media's predictive power. But one central question within this line of research is why social media offers predictive power at all given its informal and unregulated nature.

Several explanations evolved from different perspectives. Wasko and Faraj (2005) proposed that many emotional factors such as reputation, enjoyment of helping, tenure in the field, and reciprocity motivates people to contribute knowledge in social networks. Besides the emotional motivations, economic reasons also exist. Message board viewers'

reading and trading can have price impact and expedite the convergence of market prices to what the authors perceived to be fair. Because informed investors may not have the financial power to reap all the value conveyed in their private information, they have to stimulate other investors to move the market to the desired direction (Gray and Kern 2011). Informed traders also benefit from constructive feedbacks, complementary information, and confidence while communicating with their peers.

But even if the information transmitted through social media is valuable for prediction, why do investors trust and trade with it when there is no quality guarantee? Tumarkin and Whitelaw (2001) proposed several ways in which information shared on social media can influence readers: (1) the messages contain new information, (2) even if the messages do not contain new information, they at least provide an indication of general market sentiment, (3) traders may recognize the trading momentum and follow the buy and sell recommendations to exaggerate the effect.

Building on the current literature, our research is to advance the exiting theory by examining the role of network structure in affecting the predictive power of social media. A few theoretical works provide guidelines for our prediction. Colla and Mele (2010) demonstrated that linkages among traders raise the correlation of information endowment and trading behavior. In our case, when a social media discussion network is highly connected, the information correlation is high since messages linked to each other are largely focused on related issues. This leads to smaller information set coverage than what it would be if all messages were uncorrelated. Therefore, information network connectedness reduces the effective size of the information set and damages the overall predictive power.

Information free riding also plays a role in this scenario. According to the modeling work by Han and Yan (2003), when information acquisition cost is not negligible, there exists a unique equilibrium fraction of informed nodes, and it decreases with network connectedness. If acquiring private information is costly and the potential gain is not perceived to justify the search cost at the presence of free riding opportunities, people lose incentives to study the market and start to follow the crowd. The information free riders benefit from the information connections through enlarged information set and reduced risks for nearly no cost. A very connected information network implies severe information free riding, and we expect a reduced proportion of informed actors, diluted value-relevant information, and thus diminished overall informativeness in such networks.

Similar arguments date back to the model by Grossman and Stiglitz (1980). They conjectured that when the information is costly and there are informed traders in the network, the benefits of staying uninformed increase (Grossman and Stiglitz 1980). The public good literature also provides similar insights. When a new link is created, it improves the access to new information and decreases the incentive to contribute. Hence, overall welfare can be higher when there are structural holes in the network (Bramoullé and Kranton 2007).

The “hidden profile” effect also challenges the idea that group discussions are more informed than individual decisions. The discussions on social media among individual investors who have incomplete and biased information help to pool the scattered information together to form collective wisdom. However, discussions tend to overemphasize the consensus held by most individuals and to support dominant views in the group (Stasser and Titus 1985). If the public information held by all is

overwhelmingly prevailing before the discussion, people will tend to conform to the existing public opinions and overlook their private information during discussions (Qiu et al. 2016). The consequence is consensus opinions from an incomplete information set.

A highly connected discussion network on social media implies shared dominating public focuses, which partly crowd out potentially informative private information. The discussion network fails to effectively aggregate each individual's dispersive private information, leading to an incomplete and biased collective information set. In contrast, within a less connected network, the social pressure to conform is low, then dispersive private information has a better chance to draw enough attention and get incorporated into the collective wisdom. The result is a more complete information set pieced together from the scattered private information held by each social media participant.

In light of these considerations, we propose our first hypothesis:

Hypothesis 1: The intensity of information linkages within social media discussion networks is negatively related to social media's predictive power.

Following the investigation of information network connectedness, we next examine individual social media participants' nodal structures and the relationship with their predictive power. Specifically, we ask the following questions: (1) do socially connected individuals provide more value-relevant information, and (2) what type of connections matters in identifying informed participants? Ozsoylev and Walden (2011)

attempted similar questions and modeled a two-period large economy with socially connected agents. They proposed that the higher the number of connections an agent has in information networks, the higher the profits, due to increased information advantages. However, their assumption of undirected connections is a major drawback because the directions of social connections capture the flow of information between individuals, which is important in distinguishing between information free riders and information providers. Communications between social media participants can be viewed as the process of information transfer within the network. People who quote other existing messages acquire new information and people who are quoted in this process disseminate information. Therefore, the implications of incoming connections and the implications of outgoing connections are quite different, and it is necessary to tell them apart in our analysis.

Based on our arguments above, though information connections damage the overall network informativeness, they nevertheless suggest that the attractive targets for information free riders are the informed nodes. Attracting many incoming information connections implies the sharing of valuable information, so the central nodes with a high in-degree centrality are potential providers of value-relevant information in a discussion network. But for individuals who create many outgoing connections, it is a different story. Those participants are very likely to be the information free riders who learn from peers without taking much effort to search for new information.

Based on these considerations, we propose our second hypothesis:

Hypothesis 2: The participants with many incoming information connections in a social media discussion network are more likely to be providers of value-relevant information.

In the following paragraphs, we examine the implication of network connectedness to future trading intensity. Prior research studies have demonstrated the importance of network connectedness in information propagation. For well-connected nodes in a network, the information diffused from them travels rapidly because a wider audience will be exposed (Yoo et al. 2016). In our context, new information generated in a very cohesive network will be quickly passed through to other investors. Faster information spreading will cause more aggressive trading on the new information because investors will strive to profit from the new information before it is fully factored into the price.

Network linkages damage the monopolistic power of new information (Colla and Mele 2010), and cause public awareness and faster trading. Compared to a market without information linkages, a market with information linkages is more likely to experience a surge in trading intensity when new information arrives. So we propose our third hypothesis:

Hypothesis 3: The connectedness of social media discussion networks positively predicts trading intensity.

1.2 Data Description

1.2.1 Data from the Bitcoin Market

All our analyses are based on the prediction of Bitcoin market movement. In this section, we present a brief introduction to the Bitcoin market and the related data used in the study. In essence, Bitcoin is a decentralized peer-to-peer electronic payment platform. It is a web-based system that enables users to transfer values across the globe quickly and anonymously without the need for third-party verification. Though this technology resembles other electronic payment methods such as credit cards, there are a few fundamental differences: (1) Bitcoin system has underlying digital units of exchange called Bitcoins, and the exchange rates between Bitcoins and fiat currencies are decided at specialized exchanges; (2) there is no central authority maintaining the operations, regulating the issuance of currency, or keeping detailed records of transactions; (3) the entire transaction history is public information for every node in the payment network through a distributed ledger called Blockchain.

Bitcoin has seen significant growth since it was created. The market capitalization is valued at around 14 billion US dollars, and over 10 million Bitcoin wallets have been registered as of December 2016. An increasing number of businesses have accepted Bitcoin as a payment method including many industry-leading corporations such as Microsoft, Expedia, Newegg, Tesla, Home Depot, etc.

Bitcoin is established as a competitive payment platform due to several advantages over its counterparts[1]: (1) freedom in payment: payment with Bitcoin can happen

[1] For detailed explanations, please refer to "What Are the Advantages and Disadvantages of Bitcoin", *Coin Report*, accessed Dec 27, 2016, https://coinreport.net/coin-101/advantages-and-disadvantages-of-bitcoin.

anytime and anywhere with no worry of central authority limitations, (2) control and security: transactions are completed without revealing personal information, (3) information transparency: the entire transaction history is available to every one in the payment network through public address, while personal information remains hidden, (4) very low fees: there are no fees, or very low fees when faster transaction processing is needed, (5) fewer risks for merchants: due to irreversible transactions and public transparency, merchants are able to do business where crime rate is high.

To track the Bitcoin price movement, we collect Bitcoin price data from BTC-e, a major "foreign exchange" between Bitcoin and many other fiat currencies. Similar to foreign exchange markets, the Bitcoin market is open 24 hours a day, and seven days a week. The Bitcoin prices used in the analyses are the 24:00 o'clock price on each day (the daily close price). All time stamps are based on GMT. The day t Bitcoin return is calculated as $(P_t - P_{t-1})/P_{t-1}$, where P_t is the Bitcoin price on day t.

Our data spans from 2012/12/01 to 2016/06/04, including 1,282 trading days. We choose 2012/12/01 as the start date because the Bitcoin price remained low in its early years and the market capitalization was too small to attract enough public attention. The turning point occurred at the end of year 2012 when Bitcoin quickly increased in value. Table 1 presents the descriptive statistics on Bitcoin-related variables. The radical expansion of Bitcoin market is evident in Panel A of Table 1. The market capitalization of Bitcoin grows more than ten times during the study period, averaging an increase of 10 million USD every day. The growth rate of Blockchain wallet users tells a similar story. Bitcoin market is gaining popularity rapidly in recent years.

Table 1.1 – Descriptive Statistics – Bitcoin Market

	Mean	Median	Max	Min	Std. Dev	Avg. Daily Increment	Obs.
Panel A: Currency							
Units in Circulation			15,619,300	10,511,875		3,900	1,282
Market Capitalization			$13,900,051,500	$133,338,442		$10,755,245	1,282
Blockchain Wallet Users			7,504,310	46,429		5,826	1,282
Panel B: USD Based Price and Return							
Price	328.747	302.735	1,076	12.240	5.832		1,282
Return	0.43%	0.18%	41.38%	-50.31%	5.04%		1,282
Panel C: Trading Volume							
Trading Volume in	$21,023,317	$14,727,511	$240,097,870	$185,824	$679,690	$187,431	1,282
Daily # Transactions	96,844	73,923	276,448	20,555	55,236	200	1,282

Panel B of Table 1 presents descriptive statistics for Bitcoin prices and returns. Bitcoin market is very volatile, especially in the earlier years. At an infant stage, the entire system is immature; constant revolutions, disasters, and new government regulations frequently land punches on the Bitcoin market. During our data period, the highest daily return reached 41.38%, and the most fearful plummet is -50.31%.

With the development of the Bitcoin ecosystem, price volatility decreases over time. The standard deviation of the Bitcoin return is 7.23% in 2013, 3.97% in 2014, 3.62% in 2015, and 2.37% in 2016. Panel C of Table 1 presents descriptive statistics of trading volume in dollar terms and the number of transactions.

1.2.2 Data from Bitcointalk.org

Bitcointalk.org is our primary source of social data.2 It is a leading message board for Bitcoiners to share thoughts on various Bitcoin-related topics. By the time of this writing, Bitcointalk.org has 900,919 registered users and an average daily page view of 1,269,156. It receives on average 6,582 posts each day.

There are 217 boards on Bitcointalk.org, and each is dedicated to a particular topic such as technical issues, regulations, Bitcoin minings, etc. However, many of these discussion sections are not directly related to the Bitcoin market performance. In this research, we focus on the third most frequently posted board "Speculation," where people

[2] Bitcointalk.org is the only Bitcoin social media site listed on CoinGecko.com (an influential cryptocurrency summary website). On another similar website (Coinmarketcap.com), Bitcointalk.org is also listed as one of the Bitcoin message boards. The other one listed (forum.bitcoin.com) serves similar purposes and adopts a similar website layout, but there are much fewer posts. On the "official" Bitcoin website (Bitcoin.org). Bitcointalk.org is also mentioned in the "Forums" section together with the other two (Bitcoin's Reddit community and Bitcoin's StackExchange community). However, the other two are not specialized Bitcoin forums and all Bitcoin-related discussions are pooled together including topics unrelated to price discovery. Considering these factors, we choose to collect social data from Bitcointalk.org.

explicitly talk about Bitcoin price movements. From 2012/12/01 to 2016/06/04, there are over 330,000 messages written by 14,052 unique authors posted on 10,663 different threads on the speculation message board. The top two boards in terms of message count are Altcoin Announcement and Altcoin Discussions. Most Altcoins (other non-Bitcoin cryptocurrency) are very minor and there are hundreds of them in the market. These two Altcoin-related boards pool discussions regarding all non-Bitcoin crytocurrencies, so there is a huge amount of threads being created. The Speculation discussion board is actually the largest Bitcoin-related board on this forum.

Every registered user can start a new thread. After the creation of a new thread, other users can join the discussion by sharing their views in this thread. There is a communication enabling feature called "quote" on Bitcointalk.org. Users can quote one or multiple existing posts when writing a post, and a link to each quoted message is added. We construct the communication networks based on the quoting activity. Two different kinds of networks are constructed to facilitate the test of our hypotheses. Details are presented shortly after.

1.2.3 Measuring Social Media Sentiment

We follow the literature and quantify the sentiment expressed in the communications by calculating the percentage of negative words in the messages (e.g., Chen et al. 2014; Loughran and McDonald 2011; Tetlock 2007; Tetlock et al. 2008). In early studies, General Inquirer's Harvard-IV-4 classification dictionary (Harvard-IV-4 TagNeg) is used to identify the occurrence of negative words. However, Loughran and McDonald (2011) argued that the Harvard-IV-4 TagNeg substantially misclassifies words

when gauging tone in financial applications and created a new list of negative words that typically have negative implications in a financial context. We adopt the word list developed by Loughran and McDonald (2011) in our study to identify negative words. The sentiment of a discussion network in a day is calculated as the ratio of the total number of negative words to the total number of words in all related posts. To mitigate the influence of outliers that have a large ratio due to a small number of words, we winsorize the sentiment measure at the 99th percentile.

1.2.4 Social Media Discussion Network: Thread-day Networks

We collected discussions from 10,663 threads from the speculation discussion board. To test our first hypothesis, we constructed the thread-day networks (i.e., separate networks are created for each thread-day pair). Thread-day networks are used due to two considerations: (1) different threads are likely to focus on different topics, and each topic has a unique impact on future price movements, therefore it is reasonable to examine them separately, and (2) financial information is very time-sensitive, the focus and value of even the same topic may vary significantly over different days. Based on our thread-day network construction, we take two steps to test our Hypothesis 1. We first compare the predictive power of messages with and without outgoing information linkages (i.e., if the author of a message quotes other existing messages or not). Then we go deeper and measure the intensity of information linkage for each network and make comparisons.

Within each thread-day network, if node A quotes node B, a tie is created directing from A to B. A node can launch multiple outgoing information connections or receive multiple incoming information connections at the same time. It is noteworthy that in the

thread-day networks, each post, rather than each author is treated as a node. This choice is due to several considerations. First, strictly speaking, each post carries a unique piece of information, even if they are written by the same author. Discussions take place when a post is being quoted. The thread-day network with each post as a node captures the connections between each piece of information at the finest level of granularity. If an author posts uncorrelated information with different focuses in a thread-day network, treating the author as one node will bias the information connectedness measure. However, the severity of this bias depends on how correlated the author's messages are. Second, the connections between authors are not complete within a thread-day network, because the discussions between authors can go beyond a particular thread (a same group of authors can engage in discussions in multiple threads in one day). If this is the case, the connections between authors within a thread-day network is only a small segment of the bigger picture, and the connectedness of the thread-day network with authors as the nodes fails to capture the true density of connections. However, it is not possible to quote posts across threads, so the connections between posts stay within a thread and are also complete.

The downside of using posts as nodes is the flipside of the concern mentioned above. If an author posts highly correlated information within a thread-day network, the posts can be viewed as identical. In this case, treating each post as a node is equivalent to adding duplicate nodes in the network, causing biased information connectedness measure as well. Another drawback of the "post as node" construction is sparse information connections. An author having multiple posts is represented by multiple

nodes, which split the connections associated with this author. We are then faced with more sparsely connected networks than in the case of using authors as nodes.

The conclusion is that neither of the two approaches (post as node or author as node) is perfect, the decision depends on the research contexts. In order to capture information at a fine level, we choose to treat each post as a node in the thread-day network construction.

A total of 40,685 thread-day networks are constructed. Panel A of Table 2 presents descriptive statistics for all the threads. Thread Duration is the time span between the first post and the last post within a thread. The distribution of Thread Duration is highly skewed. Except for a few threads that remain active for a long time, investors quickly lose interests in a thread and the discussion desists. "% 1st Day Post" is the percentage of the first day posts. Our data indicates that on average, 69.09% of all discussions are posted within the first day. Chen et al. (2014) reports a similar pattern that the comments posted in the first two days comprise roughly 80% of all comments posted on Seeking Alpha.

Panel B of Table 2 presents descriptive statistics for the thread-day networks. PostCount is the total number of messages in a thread-day network. Many thread-day networks are very small. Later our analyses show that networks receiving only a few posts are less likely to be value-relevant. Receiving a small number of posts is a signal that this topic is not appealing to investors, or it is a redundant topic already discussed elsewhere. Furthermore, constructing network structure measures for small networks is also less meaningful. For these reasons, we primarily focus on thread-day networks with

Table 1.2 – Descriptive Statistics – Networks

	Mean	25th Percentile	Median	75th Percentile	Max	Std. Dev	Obs
Panel A: Thread							
Thread Duration (Days)	18.423	1	2	5	1,209	72.116	10,663
% 1st Day Posts	69.09%	41.67%	78.95%	100%	100%	32.47%	10,663
Daily # Threads	260.71	143	219	313.25	2,235	5.54	1,282
Panel B: Thread-day Networks							
Sentiment (% Negative Words)	1.41%	0%	1.21%	2.04%	7.14%	1.79%	40,685
Post Count	8.170	2	5	10	208	10.150	40,685
Total # Authors	6.186	2	4	8	81	6.349	40,685
Average Degree	0.499	0.250	0.500	0.732	3	0.345	40,685
Density	0.159	0.028	0.071	0.167	1	0.229	40,685

Table 1.2 (Continued)

	Mean	25th Percentile	Median	75th Percentile	Max	Std. Dev	Obs
Panel C: Thread-day Networks (PostCount>mean)							
Sentiment (% Negative Words)	1.55%	1.00%	1.46%	1.99%	6.56%	0.802%	12,507
Post Count	18.986	11	15	22	208	12.465	12,507
Total # Authors	13.285	9	11	16	81	7.069	12,507
Average Degree	0.542	0.405	0.545	0.667	1.586	0.197	12,507
Density	0.039	0.021	0.033	0.055	0.167	0.024	12,507
Panel D: Author-day Networks (PostCount>mean)							
Sentiment (% Negative Words)	1.54%	0%	0%	2.22%	12%	3.04%	104,475
In-degree Centrality	1.009	0	0	1	92	1.855	104,475
Out-degree Centrality	1.185	0	1	1	87	2.088	104,475
Author Post Count	2.619	1	2	3	120	3.107	104,475

PostCount greater than its mean (8.17) in this research. Average Degree is operationalized as the total number of incoming (or outgoing) quotations generated by all posts in a thread-day network divided by the post count. Density is operationalized as the total number of incoming (or outgoing) quotations divided by PostCount×(PostCount-1). In Panel C of Table 2, we present descriptive statistics for the thread-day networks with a post count greater than the mean that are analyzed in our hypothesis testing.

1.2.5 Social Media Discussion Network: Author-day Networks

To test our Hypothesis 2 that social media participants with many incoming information connections in a discussion network are more likely to be providers of value-relevant information, we construct daily discussion networks with each author as a node because this hypothesis requires an analysis of all posts written by each author. Since two authors can engage in discussions under multiple threads, we first break down the boundary of each thread to form 1,282 daily networks in order to characterize the connectedness between authors on each day. For each author, we then calculate the daily aggregate sentiment of all messages and the total number of incoming information connections (In-degree Centrality) in order to test Hypothesis 2.

Panel D of Table 2 presents descriptive statistics for the author-day networks. The mean of In-degree Centrality is 1.009 and the third quartile is 1, suggesting that an average participating author receives approximately one quotation each day, and that only a small fraction of authors receive many quotations. These highly quoted authors are expected to be the providers of value-relevant information. In addition, we also construct an Out-degree Centrality measure at the author/day level, which is the total number of

outgoing information connections, to examine the predictive power of social media participants who frequently quote others.

1.3 Empirical Analysis

1.3.1. Predictive Power of Social Media Discussions

Our first inference is regarding the relationship between the connectedness of social media discussions networks and the predictive power. Before introducing network structure into our analyses, it is necessary to ensure that social media sentiment predicts next-day Bitcoin return. Using the thread-day panel data, we regress the next-day return on the sentiment measure and other control variables. The baseline analysis is conducted using the following model:

$$R_{t+1}=\alpha+\alpha_t+\beta_1 Sentiment_{it}+\delta X+\eta_{it} \quad (1)$$

The dependent variable is the next-day Bitcoin return Rt+1, $Sentiment_{it}$ is the aggregate sentiment extracted from discussion thread i at time t. The coefficient estimate for $Sentiment_{it}$ reflects the effect of social media sentiment on the next-day return. The time dummy α_t (weekly dummy) controls for the differences in the returns in different time periods. X contains the intraday return R_t, the one-day lagged return R_{t-1}, the two-day lagged return R_{t-2}, the cumulative return over the past calendar month $R_{t-30,t-3}$, the logarithm of the intraday trading volume $\text{Log(TradingVolume)}_t$, and the logarithm of the number of posts Log(PostCount)_t.

We estimate a random effects model with clustered standard error (clustering by thread). $\eta_{it}=\varepsilon_{it}+u_i$, where ε_{it} is the regular disturbance and u_i is the disturbance specific to thread i. We assume that $E(\varepsilon_{it}^2|X)=\sigma_\varepsilon^2$ and $E(u_i^2|X)=\sigma_{u_i}^2$.

A random effects model is chosen over a fixed effects model because the unobserved disturbance for each thread is more probable to be random rather than fixed in different time periods. Our choice is based on the following two observations. First, participants of the same thread on different days keep changing. Every day, some new authors join the discussions and some old authors leave. This leads to fast-changing dynamics in participating members and their collective wisdom as well. Second, the topic focus of the same thread also changes with time. As new information emerges, discussions also evolve and move from one topic to another. As a result, the unobserved impact of the thread on price movement must also be changing over time. That explains why we cannot represent this unobserved impact with a fixed value.

We also conduct Hausman tests under different PostCount thresholds to compare the random and fixed effects models (Hausman 1978). The test results suggest that a random effects model is more suitable for subsamples on relatively larger networks (when PostCount > 15 and PostCount > 20) and a fixed effects model is more suitable on the full sample that includes a lot of small-size networks. This implies that as networks become larger, a random effects model is more likely to be able to capture the increasing dynamics over time than a fixed effects model.

The estimation result of Equation (1) is shown in Table 3. The coefficient estimate for Sentimentit in Column (1) is negative but statistically significant only at the 10%

Table 1.3 – Predictive Power of Social Media Discussion Network

	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)
PostCount Threshold	>Median (Median = 5.00)	> Mean (Mean = 8.17)	>15	>20
$Sentiment_{it}$	-0.105*	-0.206**	-0.453***	-0.836***
	(-1.66)	(-2.27)	(-2.80)	(-3.55)
R_t	-0.509***	-0.502***	-0.528***	-0.522***
	(-32.67)	(-26.92)	(-23.86)	(-19.02)
R_{t-1}	-0.537***	-0.526***	-0.543***	-0.566***
	(-29.27)	(-24.96)	(-19.40)	(-16.28)
R_{t-2}	-0.418 ***	-0.424***	-0.424***	-0.432***
	(-27.12)	(-23.02)	(-16.85)	(-14.27)
$R_{t-30,t-3}$	-0.133***	-0.133***	-0.135***	-0.128***
	(-19.51)	(-15.94)	(-11.79)	(-9.57)

Table 1.3 (Continued)

	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)
PostCount Threshold	>Median (Median = 5.00)	> Mean (Mean = 8.17)	>15	>20
Log(TradingVolume)$_t$	-0.033***	-0.051***	-0.058**	-0.030
	(-2.79)	(-3.34)	(-2.41)	(-1.00)
Log(PostCount)$_{it}$	0.0002	0.002	0.003	0.006
	(0.22)	(1.23)	(1.02)	(1.40)
WeekDummy	√	√	√	√
# Obs.	18,263	12,507	5,994	3,816
R^2	0.328	0.325	0.356	0.377

Note: *** p<0.01, ** p<0.05, * p<0.1.

level. The same coefficient estimates in Columns (2) to (4) are all negative and statistically significant at least at the 5% level. These results confirm the predictive power of social data sentiment for the next-day return, but only when the network is relatively large or the PostCount threshold is greater than or equal to its mean. This is expected because when a thread receives very few posts, the discussion topics are very likely not value-relevant or interesting to other investors. Going forward in Sections 4.2 to 4.4, we conduct analyses mainly based on subsamples that only contain networks with the number of posts larger than the mean.

The coefficient estimate for Sentimentit in Column (2) of Table 3 indicates that if the percentage of negative words in a thread is 1% higher, the next-day Bitcoin return is 0.206% lower. Columns (3) and (4) of Table 3 show that when the PostCount threshold becomes higher, the association between social data sentiment and next-day Bitcoin return is stronger in both magnitude (0.453% and 0.836% when PostCount > 15 and 20, respectively) and significance level (both at the 1% level). This pattern is in line with our argument that value-relevant information is concentrated in large networks. The coefficient estimates for the lagged return control variables are all negative, suggesting the presence of return reversal.

1.3.2 Network Structure and Predictive Power - Network Level Evidence

After validating the predictive power of social media sentiments, we proceed to the next step: testing the impact of discussion network structure on the predictive power. In the following subsections, we first examine the influence of information linkages at the network level, and then conduct an additional analysis at the nodal level to evaluate the

predictive power differences for messages with and without outgoing information linkages.

To test our Hypothesis 1, we compare the predictive power of discussion networks with different levels of cohesion and evaluate whether less cohesive networks are better in predicting future Bitcoin returns. For this purpose, an interaction term between sentiment and network cohesion is added to Equation (1):

$$R_{t+1}=\alpha+\alpha_t+\beta_1 Sentiment_{it}+\beta_2 Sentiment_{it}\times Cohesion_{it}+ \beta_3 Cohesion_{it}+\delta X+\eta_{it}. \quad (2)$$

Results are reported in Table 4. Two different measures are used to capture the level of network connectedness: average degree (Columns 1 to 3) and density (Columns 4 to 6). Both measures are widely-used network cohesion measures, but the average degree is more "immune to" the network size compared to the density. In our thread-day network setting, the number of outgoing information connections (the quoting) for each node is very limited (most messages only quote one other existing message, and very few quote two or more), so the network density measure decreases quickly as the network size grows. The problem lessens for the average degree measure.

Because sentiments are measured using the percentage of negative words, the predictive power is reflected in the negative coefficient estimates associated with the sentiment measure. If the intensity of information linkage damages the overall network predictive power, high network cohesion measures will etch away this predictive power

Table 1.4 – Network Structure and Predictive Power

	Cohesion Measured by Average Degree			Cohesion Measured by Density		
	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)	(5)	(6)
PostCount	> Mean	>15	>20	> Mean	>15	>20
$Sentiment_{it}$	-0.525**	-1.568***	-2.353***	-0.526***	-1.364***	-2.011***
	(-2.24)	(-3.31)	(-3.48)	(-3.01)	(-3.72)	(-3.70)
$Cohesion_{it}$	-0.0008	-0.028**	-0.033*	0.014	-0.463**	-0.635*
	(-0.12)	(-2.17)	(-1.72)	(0.25)	(-2.32)	(-1.55)
$Sentiment_{it} \times Cohesion_{it}$	0.581	2.050**	2.773**	7.016**	35.204***	57.373**
	(1.46)	(2.56)	(2.38)	(2.25)	(2.98)	(2.34)
R_t	-0.502***	-0.527***	-0.521***	-0.502***	-0.528***	-0.522***
	(-26.95)	(-23.89)	(-19.08)	(-26.99)	(-23.92)	(-19.08)
R_{t-1}	-0.527***	-0.543***	-0.566***	-0.528***	-0.543***	-0.566***
	(-25.02)	(-19.43)	(-16.38)	(-25.06)	(-19.42)	(-16.37)

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 1.4 (Continued)

	Cohesion Measured by Average Degree			Cohesion Measured by Density		
	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)	(5)	(6)
PostCount	> Mean	>15	>20	> Mean	>15	>20
R_{t-2}	-0.425***	-0.424***	-0.432***	-0.425***	-0.424***	-0.433***
	(-23.06)	(-16.89)	(-14.29)	(-23.08)	(-16.92)	(-14.33)
$R_{t-30,t-3}$	-0.133***	-0.135***	-0.127***	-0.133***	-0.135***	-0.127***
	(-15.97)	(-11.81)	(-9.58)	(-15.99)	(-11.80)	(-9.57)
$Log(TradingVolume)_t$	-0.050***	-0.058**	-0.029	-0.050***	-0.059**	-0.030
	(-3.30)	(-2.40)	(-0.95)	(-3.32)	(-2.44)	(-1.00)
$Log(PostCount)_{it}$	0.002	0.002	0.005	0.006***	0.005	0.011**
	(1.14)	(0.88)	(1.21)	(3.22)	(1.45)	(2.25)
WeekDummy	√	√	√	√	√	√
# Obs.	12,507	5,994	3,816	12,507	5,994	3,816
R^2	0.325	0.357	0.378	0.326	0.357	0.378

and dampen the negative impact of the sentiment. We expect a positive coefficient estimate for the interaction term Sentimentit × Cohesionit, if Hypothesis 1 holds true.

Columns (1) to (3) of Table 4 present the results when cohesion is measured by average degree, and Columns (4) to (6) of Table 4 present the results when cohesion is measured by density. Except for Column (1), the coefficient estimates for the interaction term Sentimentit × Cohesionit. in all other columns are positive and statistically significant at least at the 5% level. Column (2) indicates that when the percentage of negative words in discussion networks with an average degree of 0.551 (sample mean) is 1% higher, the next-day Bitcoin return is 0.438% lower (-0.438% = (-1.568+2.050*0.551)×1%). In contrast, when network cohesion is one standard deviation (0.179) higher, a 1% increase in the percentage of negative words is associated with a decrease of only 0.072% (-0.072% = (-1.568+2.050×(0.551+0.179))×1%) in the next-day Bitcoin return. In other words, the impact of social media sentiment on future returns quickly diminishes as the network becomes more cohesive. We obtain similar results when cohesion is measured by density as shown in column (5). The association between a 1% increase in the percentage of negative words and the next-day Bitcoin return is -0.504% (-0.504% = (-1.364+35.204*0.024)×1%) when the network density is at the sample mean of 0.024, but the association decreases to 0.097% (-0.097%= (-1.364+35.204*(0.024+0.012))×1%) when the network density is one standard deviation (0.012) higher. When the PostCount threshold increases, our results become stronger in Columns (3) and (6), as the coefficient estimates for the interaction term are larger in magnitude. These results support our first hypothesis that the network connectedness negatively affects the prediction accuracy of social media discussion networks.

1.3.3 Network Structure and Predictive Power - Nodal Level Evidence

In this subsection, we provide additional support for Hypothesis 1 by testing the impact of information linkages on the predictive power from a nodal perspective. Specifically, we compare between two types of messages: broadcast and discussion. Broadcasts are the standalone messages without any outgoing information connections (the messages are posted without quoting other existing messages), while discussions are the messages with outgoing information connections (the messages quote other existing messages when posted).

We examine the discussion messages first. If post A quotes other posts when it is posted, it acquires information from the quoted posts, causing information free riding. In this situation, no matter whether A is a comment or a supplement to the quoted information; it is correlated (either positively or negatively) and affiliated with the existing information set. As a result, discussion messages tend to reflect an incomplete "mirror image" of the quoted messages.

In contrast, standalone messages without quoting others (i.e., broadcasts) are expected to be less correlated with the existing discussions. They are more likely to be written by authors who want to share new information or investment opinions not yet mentioned. Also, because they are not affiliated with other messages, their arguments tend to be complete. Based on these considerations, we expect a post to be more informative when it does not originate an outgoing information linkage (a broadcast) than when it does (a discussion).

We separately calculate the sentiments of broadcasts and discussions within each thread-day network and compare their predictive power. The model specification is as follows:

$$R_{t+1}=\alpha+\alpha_t+\beta_1 Sentiment\text{-}Broadcast_{it}+\beta_2 Sentiment\text{-}Discussion_{it}+\delta X+\eta_{it} \quad (3)$$

Results are presented in Table 5. As shown in the first two rows, under different PostCount thresholds, the sentiment of the broadcasts consistently outperforms the sentiment of discussions in predicting the next-day price movements. The coefficient estimates for Sentiment-Broadcast$_{it}$ are negative and statistically significant at different levels, while the coefficient estimates for Sentiment-Discussion$_{it}$ are insignificantly different from zero. Consistent with the analysis in Secton 4.1, the broadcast messages also provide better predictive power for larger networks. The comparison between the two types of information provides additional evidence for our Hypothesis 1 that information linkages detriment network informativeness.

It is important to note that the number of observations is slightly different for the corresponding models with the same PostCount thresholds between Table 4 and Table 5. The reason is that a few threads do not have discussions. To compare between broadcast and discussion, we focus on the threads that contain both discussion and broadcast messages.

Though the results above show that the broadcast messages provide better predictions for the Bitcoin market, the fact that the first post in a thread is always a broadcast message raises the concern that the superior predictive power of the broadcast

Table 1.5 – Comparison between Broadcast and Discussion

	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PostCount Threshold	> Mean (8.17)	>15	>20	> Mean (8.17)	>15	>20	> Mean (8.17)	>15	>20
Sentiment-Broadcast$_{it}$	-0.099*	-0.229**	-0.546***				-0.094*	-0.211*	-0.544***
	(-1.84)	(-2.07)	(-3.43)				(-1.74)	(-1.87)	(-3.34)
Sentiment-Discussion$_{it}$				-0.059	-0.132	-0.148	-0.046	-0.082	-0.007
				(-1.06)	(-0.96)	(-0.69)	(-0.82)	(-0.59)	(-0.03)
R_t	-0.501***	-0.525***	-0.520***	-0.500***	-0.524***	-0.517***	-0.501***	-0.525***	-0.520***
	(-26.85)	(-23.69)	(-19.02)	(-26.80)	(-23.59)	(-18.81)	(-26.86)	(-23.69)	(-19.02)
R_{t-1}	-0.526***	-0.543***	-0.565***	-0.526***	-0.542***	-0.563***	-0.526***	-0.543***	-0.565***
	(-24.79)	(-19.34)	(-16.30)	(-24.79)	(-19.32)	(-16.18)	(-24.79)	(-19.35)	(-16.30)
R_{t-2}	-0.425***	-0.423***	-0.431***	-0.424***	-0.422***	-0.430***	-0.425***	-0.423***	-0.431***
	(-22.83)	(-16.82)	(-14.26)	(-22.79)	(-16.80)	(-14.17)	(-22.84)	(-16.83)	(-14.26)

Table 1.5 (Continued)

	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PostCount Threshold	> Mean (8.17)	>15	>20	> Mean (8.17)	>15	>20	> Mean (8.17)	>15	>20
$R_{t-30,t-3}$	-0.133*** (-15.82)	-0.135*** (-11.80)	-0.128*** (-9.55)	-0.133*** (-15.81)	-0.135*** (-11.74)	-0.128*** (-9.49)	-0.133*** (-15.82)	-0.135*** (-11.77)	-0.128*** (-9.54)
Log(Trading Volume)$_t$	-0.050*** (-3.29)	-0.057** (-2.35)	-0.024 (-0.81)	-0.051*** (-3.30)	-0.058** (-2.39)	-0.027 (-0.90)	0.050*** (-3.30)	-0.057** (-2.36)	-0.024 (-0.82)
Log(PostCount)$_{it}$	0.002 (1.08)	0.003 (1.00)	0.005 (1.33)	0.001 (1.07)	0.003 (0.95)	0.005 (1.30)	0.002 (1.09)	0.003 (1.01)	0.005 (1.33)
WeekDummy	√	√	√	√	√	√	√	√	√
# Obs.	12,388	5,985	3,815	12,388	5,985	3,815	12,388	5,985	3,815
R^2	0.324	0.355	0.376	0.324	0.355	0.374	0.324	0.356	0.376

Notes: (1) *** p<0.01, ** p<0.05, * p<0.1; (2) # Obs. are slightly smaller than those reported in Table 4 because a few threads containing only broadcast posts are dropped from analysis.

messages may come mainly from the first posts, since the messages that start a thread might be more informative. To alleviate this concern, we calculate the sentiment of broadcast messages excluding the first posts and redo the analysis. We obtain a similar result for broadcast messages' superior predictive power.

1.3.4. Author Centrality and Predictive Power

In the previous analyses, we document that the existence of information linkages diminishes the network's predictive power. However, if we view information linkages from another perspective, they also potentially serve as landmarks for informed social media actors. Authors attracting quotations from many other social media participants are likely to have shared valuable information. We expect the authors receiving many incoming information connections (i.e., getting quoted frequently) to provide more valuable information than the others.

This is an author-level analysis. During each day, the connections of authors in all threads satisfying the PostCount threshold are integrated to form daily networks. This process results in an author-day panel dataset. We compute an author's indegree-centrality in daily discussion networks as the number of incoming information connections received by the author. To examine if a high in-degree centrality implies a better predictive power, we include the interaction term between the sentiment and the in-degree centrality in our model.

We also add each author's out-degree centrality in parallel with the in-degree centrality to make a comparison. Social media participants with a high out-degree centrality are those who quote others frequently. Based on our theory, these nodes

initiating many outgoing information connections are likely to be the information free riders, and their posts offer limited insights into the future price movement. We expect to see a sharp contrast between the predictive power of social media participants who attract a lot of discussions and the predictive power of potential information free riders. We organize the analyses around Equation (4).

$$\begin{aligned} R_{t+1} = \alpha + \beta_1 Sentiment_{it} + \beta_2 Sentiment_{it} \times IndegreeCentrality_{it} + \\ \beta_3 IndegreeCentrality_{it} + \beta_4 Sentiment_{it} \times OutdegreeCentrality_{it} \\ + \beta_5 OutdegreeCentrality_{it} + \delta \boldsymbol{X} + t + \eta_{it} \end{aligned} \tag{4}$$

Different from previous models, the subscript i in Equation (4) is the author index. The results are presented in Table 6. Columns (1) to (4) present the full sample estimation and Columns (5) to (8) present the "PostCount > Mean" subsample estimation. We obtain largely similar results between these two samples, implying that considering the information revealed in the threads with a few posts provides limited benefit. Column (5) presents the baseline predictive power for an average author. When the percentage of negative words is 1% higher in an author's posts, the next-day Bitcoin return is on average 0.018% lower. The predictive power of the next-day Bitcoin return in the sentiment of an author's posts is much smaller than that of the sentiment revealed in a discussion network in Table 3 (Column 2). The primary reason is that within a discussion network, each author only carries incomplete private information, but the overall network reflects collective wisdom, which is more accurate in predicting future returns.

A comparison between Column (6) and Column (7) supports our hypothesis that we can trace the social media predictive power to the authors who attract discussions

Table 1.6 – Author Centrality and Predictive Power

	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PostCount Threshold	Full Sample	Full Sample	Full Sample	Full Sample	>Mean	>Mean	>Mean	>Mean
$Sentiment_{it}$	-0.015***	-0.008	-0.014*	-0.009	-0.018**	-0.009	-0.012	-0.007
	(-2.65)	(-1.23)	(-1.86)	(-1.18)	(-2.54)	(-1.12)	(-1.36)	(-0.76)
$IndegreeCentrality_{it}$		0.0004**		0.0004**		0.0004**		0.0003
		(2.41)		(1.97)		(2.00)		(1.59)
$Sentiment_{it} \times IndegreeCentrality_{it}$		-0.016**		-0.017**		-0.018**		-0.017**
		(-2.32)		(-2.46)		(-2.29)		(-2.18)
$OutdegreeCentrality_{it}$			0.0001	0.00002			0.0002	0.0001
			(0.86)	(0.09)			(1.08)	(0.40)
$Sentiment_{it} \times OutdegreeCentrality_{it}$			-0.002	0.003			-0.010	-0.004
			(-0.30)	(0.42)			(-1.10)	(-0.47)
R_t	-0.409***	-0.409***	-0.409***	-0.409***	-0.424***	-0.424***	-0.424***	-0.424***
	(-82.14)	(-82.13)	(-82.10)	(-82.13)	(-76.37)	(-76.37)	(-76.33)	(-76.36)

Table 1.6 (Continued)

	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PostCount Threshold	Full Sample	Full Sample	Full Sample	Full Sample	>Mean	>Mean	>Mean	>Mean
R_{t-1}	-0.482***	-0.482***	-0.482***	-0.482***	-0.494***	-0.494***	-0.494***	-0.494***
	(-96.55)	(-96.56)	(-96.57)	(-96.57)	(-85.19)	(-85.21)	(-85.22)	(-85.22)
R_{t-2}	-0.387***	-0.387***	-0.387***	-0.387***	-0.408***	-0.408***	-0.408***	-0.408***
	(-81.74)	(-81.78)	(-81.77)	(-81.79)	(-74.69)	(-74.71)	(-74.71)	(-74.71)
$R_{t-30,t-3}$	-0.111***	-0.111***	-0.111***	-0.111***	-0.118***	-0.118***	-0.118***	-0.118***
	(-54.08)	(-54.08)	(-54.08)	(-54.08)	(-47.99)	(-47.99)	(-47.99)	(-47.99)
$Log(TradingVolume)_t$	-0.021***	-0.021***	-0.021***	-0.021***	-0.027***	-0.027***	-0.027***	-0.027***
	(-5.31)	(-5.33)	(5.32)	(5.33)	(-5.09)	(-5.10)	(-5.10)	(-5.10)
WeekDummy	√	√	√	√	√	√	√	√
# Obs.	147,584	147,584	147,584	147,584	104,475	104,475	104,475	104,475
R^2	0.315	0.315	0.315	0.315	0.326	0.326	0.326	0.326

Note: *** p<0.01, ** p<0.05, * p<0.1.

from other social media participants. Specifically, when we add the interaction term between sentiment and in-degree centrality in Column (6), the coefficient estimates for Sentimentit and Sentimentit×IndegreeCentralityit are -0.009 (t-statistic = -1.12) and -0.018 (t-statistic = -2.29, statistically significant at the 5% level), respectively. These suggest that when the percentage of negative words in an author's posts is 1% higher, the next-day Bitcoin return is 0.027% (=-0.009%-0.018×1.009%) lower when the author's indegree-centrality is 1.009 (sample mean), whereas the next-day Bitcoin return is 0.061% (=-0.009%-0.018×2.864%) lower when the author's indegree-centrality is 2.864 (one standard deviation higher from the sample mean). The number of quotations an author get during a day is highly skewed. For the most central authors in the network with double digit daily quotations, their predictive power for the next-day return is much more significant.

In Column (7), author's out-degree centrality proves to be not useful in identifying the informed authors, as the coefficient estimate for the interaction term Sentimentit× OutdegreeCentralityit is statistically insignificant. This implies that authors with different out-degree centrality do not differ in their predictive power. We also test the specification of including both the in-degree centrality interaction and the out-degree centrality interaction in Column (8), and the result suggests a similar pattern.

The analyses in this section support our second hypothesis. The number of incoming information connections an author receives serves as a landmark to distinguish between informed actors and uninformed actors on social media. In contrast, the number of outgoing information connections does not suggest any difference in predictive power.

1.3.5 Network Cohesion and Trading Intensity

In this section, we examine if the network connectedness of social media discussions predicts future trading intensity. According to Colla and Mele (2010), information linkage damages the monopolistic power of the information, causing it to be traded more aggressively and incorporated into the price more quickly. Therefore, we expect that dense information linkages predict increased future trading intensity.

In our analyses, trading intensity is operationalized by two measures: the logarithm of the number of Bitcoin transactions and the logarithm of the trading volume in dollar terms. We use the thread-day networks again in this section in order to perform a panel data analysis. The model specification is shown below:

$$TradingIntensity_{t+1} = \alpha + \beta_1 Cohesion_{it} + \delta X + WeekDummy + \eta_{it} \quad (5)$$

In Equation (5), cohesion is measured by average degree or density. X includes the lagged dependent variables and the lagged returns as well (because price movements may also drive up trading intensity). Results are presented in Table 7. The first four columns and the last four columns present the results based on the two different trading intensity measures, respectively.

The first and second rows in Table 7 show our main results regarding the effect of network connectedness on future trading. Column (1) and Column (2) present the results of the regressions without control variables. In Column (3), the coefficient estimate for AverageDegreeit is 0.010 and statistically significant at the 5% level, indicating that if the network average degree is one standard deviation (0.197) higher, the number of

Table 1.7 – Network Structure and Trading Intensity

	Log(*Number of Transactions*$_{t+1}$)				Log(*Trading Volume in Dollar Terms*$_{t+1}$)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PostCount Threshold	>Mean	>Mean	>Mean	>Mean	>Mean	>Mean	>Mean	>Mean
AverageDegree$_{it}$	0.015***		0.010**		0.037***		0.027**	
	(2.81)		(2.24)		(2.65)		(2.25)	
Density$_{it}$		0.156***		0.091*		0.330**		0.230*
		(2.63)		(1.88)		(2.12)		(1.70)
TradingIntensity$_{t}$			0.417***	0.416***			0.315***	0.315***
			(39.05)	(39.02)			(30.01)	(29.99)
TradingIntensity$_{t-1}$			-0.260***	-0.260***			-0.228***	-0.228***
			(-24.00)	(-24.03)			(-25.06)	(-25.07)
TradingIntensity$_{t-2}$			-0.240***	-0.240***			-0.119***	-0.119***
			(-23.94)	(-23.93)			(-9.26)	(-9.26)
R_t			0.027*	0.027**			0.339***	0.339***
			(1.95)	(1.96)			(9.28)	(9.28)

Table 1.7 (Continued)

	Log(*Number of Transactions*$_{t+1}$)				Log(*Trading Volume in Dollar Terms*$_{t+1}$)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PostCount Threshold	>Mean	>Mean	>Mean	>Mean	>Mean	>Mean	>Mean	>Mean
R_{t-1}			0.130***	0.130***			0.341***	0.342***
			(7.85)	(7.86)			(7.21)	(7.23)
R_{t-2}			0.150***	0.150***			0.740***	0.740***
			(8.20)	(8.21)			(12.25)	(12.25)
Log(PostCount)	-0.002*	0.003	-0.001	0.002	-0.011*	0.001	-0.008	-0.008
	(-1.06)	(1.01)	(-0.66)	(0.84)	(-1.88)	(0.07)	(-1.50)	(-1.50)
WeekDummy	√	√	√	√	√	√	√	√
# Obs.	12,507	12,507	12,507	12,507	12,507	12,507	12,507	12,507
R^2	0.973	0.973	0.982	0.982	0.877	0.973	0.901	0.901

Note: *** p<0.01, ** p<0.05, * p<0.1.

transactions in the next day will increase by 0.197% (=1%×0.197). In Column (4), the coefficient estimate for Densityit is 0.091 and statistically significant at the 10% level, indicating that if the network density is one standard deviation (0.024) higher, the number of transactions in the next day will increase by 0.218% (=9.1%×0.024).

We obtain similar results when the trading volume in dollar terms is the dependent variable. In Column (7), the coefficient estimate for AverageDegreeit is 0.027 and statistically significant at the 5% level, indicating that if the network average degree is one standard deviation (0.197) higher, the trading volume in dollar terms in the next day will increase by 0.532% (=2.7%×0.197). In Column (8), the coefficient estimate for Densityit is 0.230 and statistically significant at the 10% level, indicating that if the network density is one standard deviation (0.024) higher, the trading volume in dollar terms in the next day will increase by 0.552% (=23%×0.024).

There is a significant trading intensity momentum between two consecutive days as it is shown by the positive TradingIntensityt coefficient estimates across all the full models in Table 7 (Columns 3, 4, 7, and 8). High trading volume partly continues to the next day. We also detected the reveral of trading intensity through the negative coeffcient estimates for TradingIntensityt-1 and TradingIntensityt-2. Another set of control variables (lagged returns) reveals a positive relationship between past returns and future trading intensity. The trading is very active when investors see increased prices, but conversely, they tend to hold their Bitcoin and avoid trading when price drops.

1.4 Topic Modeling

To distinguish value-relevant information from noises, this study focuses on the role of network structure embedded in social media discussion networks. To test our main Hypothesis 1, we have investigated how network structure moderates the effect of social media sentiment on future Bitcoin returns. One concern is that the actual content of social media discussions may also influence Bitcoin returns and even the network structure of discussion networks as well. For instance, certain topics are more value-relevant and may have a larger effect on returns, or different topics attract different groups of discussants and thus may be associated with specific network structures. Without controlling for the discussion contents, our analyses may potentially suffer from omitted variable bias and identify a spurious relationship.

To address this concern, we adopt topic modeling techniques to directly measure the topic distributions of each thread-day network and then add them as control variables in the model. The basic idea of topic modeling in textual analysis is that documents are represented as a distribution over latent topics, and each topic is characterized by a distribution over words. We employ Latent Dirichlet Allocation (LDA) proposed by Blei et al. (2003) to identify the latent topics from the messages posted by social media discussants and use them to infer the topic weights of thread-day networks.

For this purpose, we group all posts under a thread-day network together as one document and get a set of documents as the input for LDA. To concentrate on the words that are meaningful for identifying topics, we remove stop words such as “a”, “the”, and “of” and high frequency context-specific words such as “bitcoin” and “btc”. LDA follows the assumption of “bag of words”, so the sequence of words is ignored and only the

frequency of words matters. Following the literature, we also drop the thread-day networks that have too few words.

To train a LDA model, one important parameter to be specified is the number of topics. To find the optimal number of topics, we fit the LDA model with different numbers of topics and calculate the perplexity score of a held-out test dataset using 10-fold cross validation. The perplexity score evaluates how well a LDA model estimated on a training set of documents predicts the remaining set of documents. A lower perplexity score indicates a better language model performance. We first randomly partition the data into 10 subsamples. In each fold of the cross-validation procedure, we use 90% of data for training and hold out the remaining 10% for testing. We calculate the perplexity score of LDA models with different numbers of topics ranging from 10 to 150.

Figure 1 presents the plot between the number of topics and the average perplexity score over 10 folds. As the number of topics increases, the average perplexity score keeps decreasing at first until 30 topics and then fluctuates up and down afterwards. Among the four models (with 30, 80, 120, and 140 topics, respectively) that yield a similar perplexity score, we select 30 as the number of topics in order to fit a parsimonious model. In addition, the social media discussions we anlyze are limited to topics about Bitcoin speculation, so we do not expect a great variety of topics. We further validate our choice by manually reading the most probable words in each topic and compare the interpretability of these four models with different number of topics. We find that the interpretability of the LDA model does not improve with more topics.

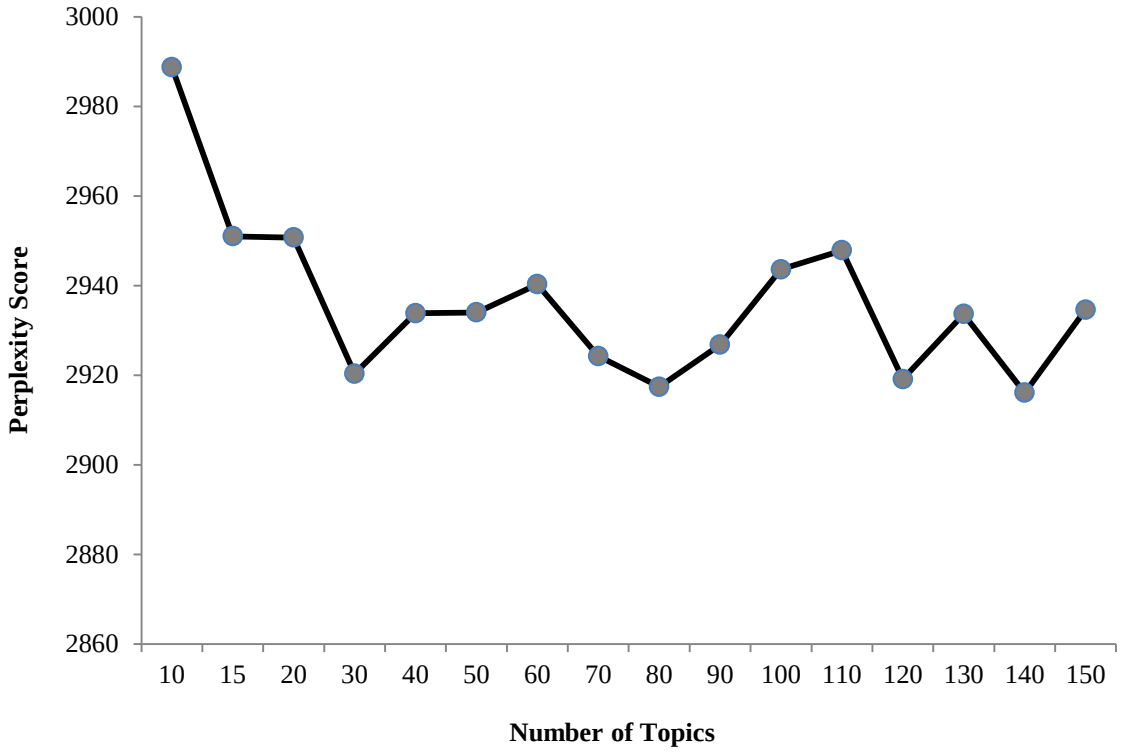

Figure 1.1 – Number of Topics and Perplexity Score

Table 1.8 – Most Probable Words in the Top 10 Topics

Topic Label	Top 20 Words
Market Risk	sell, coin, crash, drop, panic, happen, buying, make, 100, day, good, selling, big, low, money, long, 200, news, lower, cheap
Bitcoin Adoption and Application	year, make, money, world, point, early, adoption, technology, good, currency, big, future, understand, long, work, internet, investment, company, real, financial
Forum-Related Discussion	post, make, thread, money, forum, sell, coin, lose, wrong, day, troll, year, good, shit, guy, guys, stupid, lol, account, point
Summary and Trend	bubble, month, crash, year, chart, day, 2011, happen, long, rise, growth, trend, point, 2013, term, rally, increase, high, bear, news
Trading Strategy	money, make, sell, investment, lose, good, profit, risk, hold, year, long, invest, trading, term, buying, trade, day, worth, holding, month
Prospects, Wishes and Hopes	halving, rise, happen, year, good, increase, 500, month, reach, make, stable, high, block, long, hope, higher, profit, future, wait, sell
Government Regulation	government, make, power, case, road, business, world, illegal, work, fact, good, silk, state, money, law, point, bad, happen, control, free
Bitcoin Exchanges	coin, million, dollar, supply, usd, money, demand, worth, increase, number, year, cap, 100, rate, day, billion, current, inflation, exchange, fiat
Time to Reach a Price Threshold	300, 400, 500, day, month, reach, good, happen, year, rise, hope, 350, end, level, drop, pump, stable, week, range, long

Table 1.8 (Continued)

Topic Label	Top 20 Words
Price Speculation	short, trade, long, trading, make, traders, term, big, point, sell, bear, buying, money, profit, exchange, whales, volume, selling, bull, good

Table 1.9 – Network Structure and Predictive Power: Controlling for Topic Distribution

	Cohesion Measured by Average Degree			Cohesion Measured by Density		
	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)	(5)	(6)
PostCount	> Mean	>15	>20	> Mean	>15	>20
$Sentiment_{it}$	-0.486**	-1.463***	-2.248***	-0. 501***	-1.311***	-1.969***
	(-2.07)	(-3.04)	(-3.29)	(-2.81)	(-3.48)	(-3.63)
$Cohesion_{it}$	0.001	-0.026**	-0.034*	0.036	-0.446**	-0.656
	(0.21)	(-1.98)	(-1.70)	(0.62)	(-2.19)	(-1.61)
$Sentiment_{it} \times Cohesion_{it}$	0.511	1.875**	2.578**	6.431**	33.122***	54.687**
	(1.29)	(2.32)	(2.19)	(2.07)	(2.81)	(2.24)
R_t	-0.502***	-0.527***	-0.524***	-0.502***	-0.527***	-0.525***
	(-26.65)	(-23.76)	(-19.01)	(-26.69)	(-23.78)	(-19.03)
R_{t-1}	-0.529***	-0.546***	-0.572***	-0.530***	-0.546***	-0.572***
	(-25.04)	(-19.33)	(-16.38)	(-25.07)	(-19.33)	(-16.38)

Table 1.9 (Continued)

	Cohesion Measured by Average Degree			Cohesion Measured by Density		
	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)	(5)	(6)
PostCount	> Mean	>15	>20	> Mean	>15	>20
R_{t-2}	-0.425***	-0.422***	-0.434***	-0.425***	-0.422***	-0.435***
	(-23.04)	(-16.77)	(-14.32)	(-23.06)	(-16.79)	(-14.35)
$R_{t-30,t-3}$	-0.132***	-0.135***	-0.128***	-0.133***	-0.136***	-0.128***
	(-15.89)	(-11.72)	(-9.56)	(-15.92)	(-11.72)	(-9.55)
$Log(TradingVolume)_t$	-0.046***	-0.052**	-0.022	-0.046***	-0.053**	-0.023
	(-3.01)	(-2.14)	(-0.71)	(-3.02)	(-2.17)	(-0.74)
Log(PostCount)	0.002	0.003	0.005	0.007***	0.005	0.009*
	(1.33)	(0.98)	(1.08)	(3.57)	(1.39)	(1.82)
WeekDummy	√	√	√	√	√	√
Topic Distribution	√	√	√	√	√	√

Table 1.9 (Continued)

	Cohesion Measured by Average Degree			Cohesion Measured by Density		
	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)	(5)	(6)
PostCount	> Mean	>15	>20	> Mean	>15	>20
# Obs.	12,444	5,961	3,797	12,444	5,961	3,797
R^2	0.328	0.360	0.384	0.328	0.360	0.384

Note: *** p<0.01, ** p<0.05, * p<0.1.

To illustrate the topics uncovered by the LDA model, we try to provide an intuitive label for each topic based on its most probable words. Table 8 presents the topic labels we summarize and the 20 most probable words in each of the top 10 topics ranked by topic weights. Many topics share a few common keywords because the scope of disussions in our reseach context is quite limited; people are communicating about the Bitcoin price movement in one way or another. This table shows the effectiveness of the LDA model in identifiying the latent topics in a set of documents.

The LDA model also produces a topic distribution over the 30 topics for each document (thread-day discussion network). This topic distribution represents the weights of different topics in a document, and all the weights add up to 1. To control for the contents of social media discussions, we add these topic weights as additional control variables to Equation (2) and re-estimate our main model as in Table 4. The new results are presented in Table 9. The coefficient estimates on Sentimentit×Cohesionit are still statistically significant at least at the 5% level except for the first column, although the magnitude of the effects is slightly less than the corresponding results in Table 4. We conclude that our main finding remains the same after accounting for the topic contents of social media discussions and that network structure plays an important role in distinguishing value-relevant information from noises.

1.5 Conclusion

This study aims to investigate the role of network structure on the predictive power of social media for the Bitcoin market. By analyzing the discussion networks related to the Bitcoin market, we evidence a negative relationship between network connectedness

and its predictive power for future price movement. Further investigation at the individual level shows that social media participants who attract discussions are more likely to share value-relevant information. We also found that the network connectedness predict future trading intensity.

SOCIAL MEDIA INCENTIVE HIERARCHY AND USER PREDICTION ACCURACY

Incentive hierarchies are common practice in online gaming as a way to motive user activities. In recent years, many social media platforms also implemented incentive hierarchies to gamify the user experience in order to encourage participation and contribution. The fundamental idea is to help users internalize the benefits from content sharing in a "free-riding" environment where all information posted is public goods available to everyone (Goes et al. 2016). Incentive hierarchies are so widely used nowadays and almost every social media platform is using it. Examples include Stack Overflow, Foursquare, and many more.

The social media incentive hierarchy is mainly designed to achieve five purposes according to Antin and Churchill (2011): (1) Goal setting: badges are used to motivate activities; (2) Instruction: badges are used to instruct the new users and help them to diversify participation; (3) Reputation: badges are used to provide a user's interests, expertise, past interaction, and engagement level; (4) Status/Affirmation: badges are used to advertise user's past achievement without explicit bragging; (5) Group Identification: badges communicate a set of activities that bind a group of users together around shared experience.

In most cases, social media incentive hierarchy systems allow users to accumulate points for engaging in various social interactions or contributing new contents, and then award badges/trophies/levels/ranks when their cumulative points reach a threshold. Users displaying better incentive hierarchies imply that they accomplished many goals and participated in many discussions. But do these users always provide high-quality information?

This is the fundamental question we ask in this research. We compare the motivation between high rank users and low rank users, and empirically answer this question using data from a social media platform specialized in discussions about the digital currency market. Our paper mainly contributes to the related literature in analyzing the effectiveness of social media incentive hierarchy systems and in providing guidelines in devising better badge-awarding rules. We also complement the literature of social media's role in the financial markets.

To evaluate the implication of social media incentive hierarchy on users' prediction accuracy, we start from discussing social media users' motivation to share private information as well as their cost in doing so. Previous literature provided theoretical guidelines to our investigations.

Social media users incur time and effort costs to share a piece of private information with others. They have to first locate a familiar topic, read others' discussions, collect complementary information from other resources, and finally formulate and edit a response. In addition to the time and effort costs, they also forfeit their information advantage because their private information now becomes public information. Given all the costs, the sharing users must be motivated for some reasons to compensate for the costs.

Several motivations have been identified from different perspectives. Wasko and Faraj (2005) suggested a few socialization purposes and emotional factors that potentially incentivize people to communicate online with others: (1) reputation: the approval, respect, and status gained when engaging in social interaction (Blau 1964); (2) enjoying helping: the

good feelings and enjoyment when helping others (Kollock 1999); (3) individual's structural centrality increase willingness to contribute (Wasko and Faraj 2005); (4) commitment: the perceived duty and obligation to engage in interactions (Coleman 1990); (5) reciprocity: the perceived moral obligation to pay back to peers and the network (Wasko and Faraj 2000).

Besides the socialization purposes, social media users also share information for economic purposes. Informed traders may benefit from constructive feedbacks, complementary information, and confidence in trading while communicating with their peers (Gray and Kern 2011). In other situations, when informed traders have quality information but lack the financial power to correct the price discrepancy and realize the profits, they have the incentive to publicize their private information to create a trading momentum so that together with their peers, the informed traders can move the market to the desired direction and then profit from their private information.

Taken together, the sharing user has both socialization-related motivations and economic-related motivations to justify the cost associated with the sharing activities. We argue that when social media users are motivated mainly by economic purposes rather than by socialization purposes, they provide more value-relevant information.

The activity-based rank designed by Bitcointalk.org is a measure of users' level of social media activity. Only consistently active users over a long time period are granted high ranks. They frequently participate in activities such as posting and commenting. The hard-earned rank is a demonstration of their level of engagement. They are generally more social people, at least in online communities. They enjoy social awards and benefit more from

emotional comfort than others via socialization. They are also better recognized and have more relationships to maintain compared to their less active peers.

Therefore, different from inactive users with lower ranks, active social media users with high ranks have comparatively more socialization-related motivations to post. For them, the cost associated with online sharing (time, effort, and loss of information advantage) are more easily justified without significant economic purposes.

The situation is different for low rank users. Low rank users are comparatively less active in peer communications and less motivated by socialization-related reasons. They don't benefit a lot from social interaction itself. Low badge users share private information not so much for social awards, but most likely for other economic-related reasons. When they post, they are more likely to be stimulated by some new information and want to make a statement, not just to socialize. Based on our argument above, we expect these inactive users holding lower badges to be more value-relevant on average.

We also draw insights from the Drive-Reduction Theory (Dewey 2007). Basically, the motivation drops after the goal is reached. In most cases, the social media incentive hierarchies are permanently offered when they are obtained. Users will lose the incentives to upgrade their status because there is no more room to improve.

The second part of our analysis is about users' visibility and influence in the online community. The ranking systems are implemented to establish status via repetitive usage. And the ranks advertise one's achievements and past accomplishments (Antin and Churchill 2011). High rank users enjoy better visibility in the online community and are treated as

experienced seniors in a particular field when they display their advanced ranks. As a result, their posts have wider exposure to the public than social media participants with relatively lower ranks. We expect the information conveyed by higher rank users to induce a broader influence.

In contrast, low rank users, being inactive through time, usually have a lower presence in public, and are not well recognized by their peers. To make things worse, the content viewers probably are mainly attracted by high rank users and are reluctant to spend much time on low rank users' posts due to time and effort costs. As a result, fewer social media users in the online community pay attention to low rank users' posts compared to high rank users'. And we expect a smaller impact from them.

To this point, we try to raise a doubt about the effectiveness of activity-based incentive hierarchy systems. While high level users enjoy high exposures, the proportion of value-relevant information is lower due to reduced motivations. Conversely, while low rank users have better potential to devote more efforts in formulating quality posts, their ideas and shared information do not get enough attention from others. Recognizing this dilemma, we can improve the accuracy in predicting the future market movements based on social media discussions.

To carry out our analyses, we downloaded social media discussions from a leading message board in the digital currency field called Bitcointalk.org, in the period of February 2015 to February 2017. Our sample consists of more than 190,000 messages from the Bitcoin-related discussion section, and 620,000 messages from 66 Altcoin-related discussion

sections (Altcoin is the abbreviation for all non-Bitcoin cryptocurrencies). We also downloaded the historical price data for Bitcoin and all Altcoins during the same period.

We first conduct our baseline analysis by investigating the predictive power of the aggregated Altcoin discussion sentiments. Then we verify the superior predictive accuracy from the low rank users. We do so by breaking down the aggregated discussion sentiment into two parts: the high rank user sentiment and the low rank user sentiment, and compare the predictive power of the two user segments.

Next, in order to compare the visibility and influence from high rank users and low rank users, we study their spillover effect. Specifically, we test how high rank users' discussion and low rank users' discussion on the Bitcoin-related discussion board affect the next-day Altcoin returns differently. Many studies in the finance literature suggested that events occurred to a company can cause intra-industry spillover and even inter-industry spillover (Hsu et al. 2010, Helwege and Zhang 2015, Chen et al. 2005, Otchere 2007, Goins and Gruca 2008, Tawatnuntachai and D'Mello 2002, Elliott et al. 2006, Akhigbe et al. 2015). We argue that more visible and influential information shared on social media will diffuse more easily and will be more effective in causing the spillover effect.

To summarize, this research mainly contributes to the social media incentive hierarchy system literature by studying how incentive hierarchy system shapes users' motivation for contribution in online communities and suggesting methods to infer social media users' informativeness and influence based on the users' ranks obtained from social media. The literature just started analyzing the effectiveness of incentive hierarchy system recently, and

most of existing studies are focused on the quantity of online activities, but little is done to study its influence on the quality of online activities. We also contribute to the information spillover literature by studying the information spillover through social media. Most of prior research in this literature focus on a single event at a time (such as bankruptcy) and investigate if the event induces a positive spillover (contagion) or a negative spillover (competition) effect. However, in recent years, besides those major shocks that happen not so frequently, a comprehensive mixture of business information is transmitted through social media at a much higher frequency. It is thus necessary to extend the related literature to examine information spillover through social media.

Our research also provides practical implications. We demonstrate that social media incentive hierarchy systems can be used to sort out the valuable investment advice from an enormous amount of social data generated each day. Our insights will help reduce information acquisition cost.

2.1 Literature Review

Social media hierarchy system is a form of gamification applications. Gamification is defined as the use of game design elements in non-game contexts (Deterding et al. 2011), and the primary purpose is to motivate user participation and contribution. Our research is among those focusing on the effectiveness of gamification applications. Many areas have been investigated, such as education (Cheong et al. 2013, Dominguez et al. 2013, Denny 2013), Intra-organization systems (Farzan and Brusilovsky 2011, Thom et al. 2012), Q&A forums (Anderson et al. 2013, Grant and Buddy 2013, Goes et al. 2016), and Ideation (Jung et al.

2010, Witt et al. 2011). Our study is closely related to those about the Q&A forums. Most studies in this field mainly investigate the effectiveness of gamification on user activities in those forums, but little attention is paid to users' information quality, which is a more meaningful topic. The lack of such research lies in the difficulty to measure the information quality. We contribute to this line of literature by filling this cap.

The incentive hierarchies used on social media are mainly designed to achieve five purposes: goal setting, new user instructions, reputation system, status and affirmation, and group identification. However, much work remains to be done to fully understand the positive and negative influences (Antin and Churchill 2011).

This study builds on the recent finding that social media contents provide valuable insights into future return predictions in the financial markets. Actually, researchers have long been aware that traditional financial reports and editorial media can predict stock market returns (Davis et al. 2011, Loughran and McDonald 2011, Tetlock 2007, Tetlock et al. 2008, Solomon 2012). Studies also show that the discussions on many online message boards demonstrate predictive powers for the price movement, even though social media discussions are unregulated and there is no guarantee for the information quality (Tumarkin and Whitelaw 2001, Das and Chen 2007, Chen et al. 2014).

On social media platforms, the incentive hierarchy is a representation of a user's past achievements and level of participation in online activities such as posting and commenting. Intuitively, others will look more favorably upon someone who has undertaken a series of activities that earn a certain rank. But does this necessarily imply superior information quality

in terms of predicting future returns? Does the advanced badge make a user more visible and influential to others? To answer these questions, it is important to dissect the motivation to contribute from the standpoints of a high rank user and of a low rank user.

To share information or communicate with peers, social media users have to first access the network, review opinions, information, and questions posted by others, and choose the ones they feel comfortable and capable to respond, and then take time to write up the post. By posting it, the users give up their private information for free. From this standpoint, sharing private information benefits everyone else but sharers. Obviously, the cost associated with the sharing activity has to be justified. Wasko (2005) drew on prior research and theories on collective action and summarized the motivation for online sharing as Reputation, Enjoy Helping, Centrality, Self-rated Expertise Tenure in the Field, Commitment, and Reciprocity. These motivations can be characterized as socialization-related motivations.

Besides the socialization purposes, the finance literature in particular also pointed to the existence of economic purposes. Message board viewers' reading and trading can have price impact and expedite the convergence of market prices to what the sharer perceived to be fair. Because informed investors may not have the financial power to reap all the value conveyed in their private information, they have to stimulate other investors to move the market to the desired direction (Gray and Kern 2011). Informed traders also benefit from constructive feedbacks, complementary information, and confidence in trading while communicating with their peers.

Based on the analysis, we argue that the motivation for online sharing mainly consists of two parts: socialization-related motivations and economic-related motivations. Low rank users are less likely to be social people, at least in online communities. As a result, they do not benefit from social awards and emotional comfort as much as those high rank users. To compensate the cost of sharing, they must be strongly motivated by economic benefits to make a point on social media. While with more emphasis on social awards, high badge users are better motivated by socializing with people, so they are less focused on value-relevant information.

Intuitively, high rank users engage in social media activities not only for communicating value-relevant information, but also for the purpose of socialization. The result is a higher proportion of irrelevant and off-topic activities and diluted informativeness. In contrast, though low rank users engage in social media activities less frequently, when they do, they mostly likely talk business.

Our prediction is also supported by the Drive-Reduction Theory (Dewey 2007). The Drive-Reduction theory states that the motivation drops after the goal is reached. In most cases, the social media incentive hierarchies can be viewed as a set of goals and they are permanently awarded when they are obtained. As a result, users will lose the incentives to keep sharing quality contents after the ranks are awarded. Conversely, social media users who value the respect and status from a higher rank but are currently at a lower rank must have stronger incentives to share quality information. So we predict that social media users

with low ranks in hierarchy predict future price movements more accurately than social media users with high ranks in hierarchy.

Though we have shown that social media users with high ranks provide less value-relevant information than their counterparts with low ranks, they shall have wider influence among peers for the following two reasons. First, high rank users are expected to have more connections with other social media users due to their active participation in online social interactions. They are better recognized in the community. Social network theories suggested that the number of social connections plays an important role in speeding up the information diffusion because individuals who frequently interact with others are more likely to be influential (Brown and Peter 1987). Messages written by high rank users who enjoy more connections with peers will diffuse faster among the social network than messages written by low rank users.

Second, because the ranks awarded to a user reflect the user' past accomplishments and experience, and other users can use it to infer the trustworthiness and reliability of the content (Anton and Churchill 2011). Therefore, for a given message, if it is posted by a user displaying a high rank, it is more attractive than if it is posted by a low rank user. Based on the arguments above, we expect that messages posted by social media users with high ranks in hierarchy have wider influence.

In the finance literature, it is well established that the impact of new information can travel beyond the boundary of the affected firm and cause spillover effects. The spillover effect is positive when the affected firm and its rivals react in the same direction, and it is

negative when the affected firm and its rivals react in the opposite direction. Many types of events have been shown to induce spillover effects due to shared technologies, business models, resources, and customers within the same industry, such as bankruptcy (Ferris et al. 1997, Helwege and Zhang 2015, Lang and Stultz 1992), IPO announcements (Hsu et al. 2010), new product introductions (Chen et al. 2005), merger announcements (Akhigbe and Martin 2000), dividend-related announcements (Laux et al. 1998, Slovin et al. 1999), privatization announcements (Otchere 2007), layoff announcements (Goins and Gruc 2008), stock split announcements (Tawatnuntachai and D'Mello 2002), going-concern audit opinions (Elliott 2006), and stock price surprises (Akhigbe et al. 2015). Unlike those major events, the information diffusing through social media is more frequent and contains a mixture of information with different level of importance. We argue that only the most visible information potentially induces the spillover effect, and the social media users with high ranks in hierarchy are the mostly probable providers of such visible information. Thus, we predict that social media users with higher ranks induce stronger spillover effects than social media users with lower ranks.

2.2 Data Description

2.2.1 Data from the Bitcoin Market

Our analyses are conducted in the context of predicting the market movement for the cryptocurrencies. In this section, we first present a brief introduction to the Bitcoin market and the related data used in the study. Bitcoin is a decentralized peer-to-peer electronic

payment platform. It is a web-based system that enables users to transfer values across the globe quickly and anonymously without the need for third-party verification.

Bitcoin has seen significant growth since it was created. The market capitalization is valued at around 45 billion US dollars as of July 2017. An increasing number of businesses have accepted Bitcoin as a payment method including many industry-leading corporations such as Microsoft, Expedia, Newegg, Tesla, and Home Depot.

To track the Bitcoin price movement, we collect Bitcoin price data from Poloniex, a major "foreign exchange" between Bitcoin and USD. Though it is not the largest Bitcoin-USD exchange, it also provides historical price information for many Altcoins ("Altcoin" usually refers to all non-Bitcoin cryptocurrencies, short for "alternative to Bitcoin"). Similar to traditional foreign exchange markets, the cryptocurrency markets are active 24 hours a day, and seven days a week. The Bitcoin prices used in the analyses are the 24:00 o'clock price on each day (the daily close price). All time stamps are based on GMT. The day t Bitcoin return is calculated as $(P_t - P_{t-1})/P_{t-1}$, where P_t is the Bitcoin price on day t. Our data spans from 2015/2/19 to 2017/2/17. We choose 2015/2/19 as the start date because it is the earliest trading date on Poloniex. Panel A of Table 2.1 presents the descriptive statistics on Bitcoin-related variables.

2.2.2 Data from the Altcoin Market

Altcoins are all non-Bitcoin cryptocurrencies, and they use similar technologies as Bitcoin but usually adopt a different monetary policy such as currency issuance rules,

transaction confirmation methods, and mining methods. We can treat them as intra-industry competitors to Bitcoin because of the technology similarity.

Bitcoin is the only crytocurrency in market for a long time since it was first created in 2009. Starting in 2014, the development of Altcoins flourished, and all of a sudden, dozens of Altcoins emerged. While many of them soon went out of market due to extremely inactive trading, some of them survived and grew rapidly in market capitalization and public attention. Though there are over a thousand Altcoins with active trading, we limit our attention to those major competitors listed on the Poloniex exchange. Similarly, the data spans from 2015/2/19 to 2017/2/17. All time stamps are based on GMT. Panel B of Table 2.1 presents the descriptive statistics on Altcoin-related variables.

2.2.3 Bitcoin and Altcoin Social Data

We downloaded the social media discussion data from Bitcointalk.org. It is a leading message board for cryptocurrency investors to share thoughts on various topics. At the time of writing, Bitcointalk.org has 969,611 registered users and an average daily page view of 1,342,470. It receives on average 6,860.21 posts each day.

There are 217 boards on Bitcointalk.org. To collect the Bitcoin-related social discussions, we use the discussions from the “Speculation” discussion section. Though most of the 217 boards are dedicated to Bitcoin-related discussions, not all of them are as closely related to Bitcoin pricing (such as technical issues, regulations, and Bitcoin mining) as the Speculation discussion board.

Table 2.1– Descriptive Statistics – Bitcoin Market

	Mean	Median	Max	Min	Std. Dev	Obs.
Panel A: Daily Bitcoin Return						
Aggregated Daily Sentiment	0.014	0.014	0.024	0.007	0.003	730
Daily Return	0.36%	0.293%	18.66%	-31.89%	3.30%	730
Close Price	467.092	421.782	1136	178.719	218.656	730
# Post	256.893	234	1211	4	136.761	730
Panel B: Altcoin						
Aggregated Daily Sentiment	0.012	0.008	1	0	0.040	33,083
Daily Return	1.40%	-0.11%	2684.06%	-99.99%	24.07%	20,882
# Post	18.837	5	2,160	0	54.052	33,083
# Author	8.913	4	446	0	16.020	33,083
Panel C: Bitcoin Thread-Day Return						
Thread-Day Sentiment	0.013	0.010	1	0	0.016	28,194
# Post	6.882	4	232	1	9.251	28,194

The choice of the “Speculation” discussion section is to focus on the most relevant information and avoid introducing noise information in our analysis. We downloaded all 190,000 messages posted from the speculation board. Besides Bitcoin-related discussions, Bitcointalk.org also provided places for Altcoin discussions. The largest and the most popular board in terms of post volume is the “Altcoin Announcement” discussion section. It may seem ironic at first that the most popular discussion board on Bitcointalk.org is about Altcoins. This is however due to the large quantity of Altcoins being discussed. Within the “Altcoin Announcement” board, new Altcoins are announced with a new thread, and the title of the thread follows a standard format that can be used to uniquely identify the Altcoin (similarly as a ticker symbol in the stock market). We successfully located the discussion threads for the 66 actively traded Altcoins listed on Poloniex. We downloaded over 600,000 messages in total posted on these Altcoins. A comparison between Panel A and Panel B in Table 2.1 shows that the post volume for Altcoin discussions is quite small compared to Bitcoin discussions.

2.2.4 Extracting Sentiment from Social Data

We follow the literature and quantify the sentiment expressed in the communications by calculating the percentage of negative words in the messages (e.g., Chen et al. 2014; Loughran and McDonald 2011; Tetlock 2007; Tetlock et al. 2008). In early studies, General Inquirer’s Harvard-IV-4 classification dictionary (Harvard-IV-4 TagNeg) is used to identify the occurrence of negative words. However, Loughran and McDonald (2011) argued that the Harvard-IV-4 TagNeg substantially misclassifies words when gauging tone in financial

applications and created a new list of negative words that typically have negative implications in a financial context. We adopt the word list developed by Loughran and McDonald (2011) in our study to identify negative words. The sentiment of a discussion network in a day is calculated as the ratio of the total number of negative words to the total number of words in all related posts.

2.2.5 Incentive Hierarchy System in Bitcointalk.org

Bitcointalk.org employs a simple activity-based incentive hierarchy system. The purpose of introducing this system is to encourage participation. Similar incentive hierarchy systems have been deployed in many social media platforms. For Bitcointalk.org users, the formula used to calculate their activity points is $\min(\text{time} \times 14, \text{number of posts or comments})$, where time is the number of two-week periods when the user is active since registration. To get high points, the user must be (1) actively participating in discussions, and (2) remain active for a long period of time. Though the method to calculate the user points differs on different sites, the basic principle is largely the same.

A higher rank is awarded when a user's cumulative point reaches a threshold. Bitcointalk.org offers eight levels: Brand New, Newbie, Jr. Member, Member, Full Member, Sr. Member, Hero Member, and Legendary. To compare the informativeness of users with different ranks, we separately calculate the sentiment of high rank users and low rank users. Different cutoff values are used to fully explore the relationship.

One thing to note is that the ranks in our data are the ones observed at the end of the data collection period. Though it is not the rank the users were holding at the time of the post, it represents the user's willingness to engage in online social activities because high rank users must keep being active for a long time. Therefore, the ranks at the end also reflect users' motivation to engage in social exchanges.

To be robust, we also downloaded the users' entire posting history and calculated their rank at the time of each post based on the activity score formula.

2.3 Empirical Analysis

2.3.1. Incentive Hierarchy and Prediction Accuracy: Evidence from the Altcoin Markets

We test our predictions for the Altcoin market and the Bitcoin market separately for the following considerations. The Bitcoin discussion board we selected (the "Speculation" discussion board) is expected to contain the most relevant information for the Bitcoin price movement, many other discussion boards are less relevant. We only use the speculation board to be conservative so that irrelevant information would not affect our prediction. However, in the Altcoin discussion section, under each Altcoin threads, the discussions are not categorized based on the topics. All related discussions are pooled together.

In this section, we first focus on the price prediction for the 66 Altcoins in our sample. We employed a fixed effect linear model with each Altcoin as a cross section to test our first prediction. The next-day return is regressed on the sentiment measures and other control variables. The analysis is conducted using the following model:

$$R_{i,t+1}=\alpha+\beta_1 HRSentiment_{i,t}+\beta_2 LRSentiment_{i,t}+\delta \boldsymbol{X}+i+a_t+\eta_{i,t} \quad (1)$$

The dependent variable $R_{i,t+1}$ is the next-day return for Altcoin *i* on day *t*+1, $HRSentiment_{i,t}$ is the daily aggregate sentiment extracted from social media discussions posted by high rank users for Altcoin *i* on day *t*. $LRSentiment_{i,t}$ is the daily aggregate sentiment extracted from social media discussions posted by low rank users for Altcoin *i* on day *t*. The eight ranks awarded by Bitcontalk.org are Brand New, Newbie, Jr. Member, Member, Full Member, Sr. Member, Hero Member, and Legendary. The high rank user threshold we use is Full Member or above (120 activity points or more). Alternative thresholds are also tested. The coefficient estimates for the two sentiment measures reflect the effect of social media sentiment on the next-day return. The time dummy α_t (week dummy) controls for the differences in the returns in different time periods. The Altcoin dummy *i* controls for the altcoin specific fixed effect. *X* contains the time *t* return for Altcoin *i* and Bitcoin: $ALTR_{i,t}$ and $BTCR_{i,t}$, the one-day lagged return for Altcoin *i* and Bitcoin: $\mathrm{ALTR}_{i,t-1}$ and $\mathrm{BTCR}_{i,t-1}$, the two-day lagged return for Altcoin *i* and Bitcoin: $ALTR_{i,t-2}$ and $BTCR_{i,t-2}$, the logarithm of the time *t* post count for Altcoin *i* $Log(AltPostCount)_{i,t}$, the logarithm of the time *t* author count for Altcoin *i* $Log(AltAuthorCount)_{i,t}$, and weekly market capitalization share for Altcoin *i* $MarketCapShare_{i,t}$ ranging from 0 to 1.

We conducted Hausman test to verify our choice of the fixed effect model, but the Hausman test result does not reject the use of random effect model. Therefore, we present both the fixed effect model result and the random effect model result. The estimation of Equation (1) is shown in Table 2.2, Column (1) to Column (5).

The coefficient estimate for $TotalSentiment_{i,t}$ in Column (1) is negative but statistically insignificant, indicating that overall the information is noisy. In Column (2) and Column (3), we breakdown the total sentiment into $HRSentiment_{i,t}$ and $LRSentiment_{i,t}$, and include them in the regression separately. The results show a stronger predictive power from $LRSentiment_{i,t}$, the sentiment of low rank users. The coefficient estimate on $LRSentiment_{i,t}$ in Column (3) is -0.149 and significant at the 5% level, meaning that if the overall percentage of negative words in the Altcoin discussion is 1% higher, the next-day return for that Altcoin will be lower by 0.149%.

Column (4) of Table 2.2 presents the result when we include both $HRSentiment_{i,t}$ and $LRSentiment_{i,t}$ in the regression, and the coefficient on $LRSentiment_{i,t}$ is -0.155 and significant at the 5% level, which is similar to the result in Column (2). In Column (5) of Table 2.2, we use the social media users' rank at the time of each post, and a new set of $HRSentiment_{i,t}$ and $LRSentiment_{i,t}$ are calculated based on the new ranks. The coefficient on $LRSentiment_{i,t}$ becomes even more negative (-0.190) and still significant at the 5% level. In Column (6), we estimate the same model using random effect estimation with clustered standard error (as the Hausman test indicates that the random effect model is appropriate), and the coefficient estimate on $LRSentiment_{i,t}$ is -0.192 at the 1% significant level. In contrast, the coefficient estimates on $HRSentiment_{i,t}$ are not statistically significant across all model specifications (Column 1 through Column 6). These results support our first hypothesis that the low rank social media users provide better prediction for future price movements.

Table 2.2 – Predictive Power of Social Media Users with Different Incentive Hierarchy Rank: Altcoin

	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$
	(1)	(2)	(3)	(4)	(5)	(6)
	TotalSentiment	HRSentiment	LRSentiment	Both Sentiments	Rank at Post	Rank at Post, Random Effect
$TotalSentiment_{i,t}$	0.006					
	(0.13)					
$HRSentiment_{i,t}$		0.165		0.177	0.014	0.0003
		(1.40)		(1.50)	(0.14)	(0.00)
$LRSentiment_{i,t}$			-0.149**	-0.155**	-0.190**	-0.192***
			(-1.99)	(-2.06)	(-2.04)	(-2.88)
$ALTR_{i,t}$	-0.037***	-0.027***	-0.028***	-0.028***	-0.026**	-0.020
	(-5.44)	(-3.14)	(-3.21)	(-3.17)	(-2.54)	(-1.06)
$ALTR_{i,t-1}$	-0.028***	-0.018***	-0.018***	-0.018***	-0.018***	-0.017*
	(-4.03)	(-3.99)	(-4.01)	(-4.00)	(-3.68)	(-1.67)
$ALTR_{i,t-2}$	-0.013	-0.014***	-0.014***	-0.014***	-0.011**	-0.010***
	(-1.84)	(-2.88)	(-2.89)	(-2.88)	(-2.16)	(-4.67)

Table 2.2 (Continued)

	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$
	(1)	(2)	(3)	(4)	(5)	(6)
	TotalSentiment	HRSentiment	LRSentiment	Both Sentiments	Rank at Post	Rank at Post, Random Effect
$BTCR_{i,t}$	-0.303***	-0.312***	-0.315***	-0.314***	-0.330***	-0.335***
	(-5.42)	(-5.84)	(-5.90)	(-5.88)	(-5.03)	(-7.25)
$BTCR_{i,t-1}$	-0.169***	-0.192***	-0.192***	-0.193***	-0.205***	-0.198***
	(-2.98)	(-3.60)	(-3.61)	(-3.64)	(-3.14)	(-3.94)
$BTCR_{i,t-2}$	-0.141**	-0.167***	-0.166***	-0.167***	-0.190***	-0.193***
	(-2.49)	(-3.14)	(-3.11)	(-3.14)	(-2.94)	(-3.10)
$Log(PostCount_{i,t})$	0.012*	0.0001	0.001	0.0004	-0.002	0.002
	(1.82)	(0.02)	(0.11)	(0.06)	(-0.28)	(0.35)
$Log(AuthorCount_{i,t})$	-0.014*	-0.003	-0.004	-0.003	0.0005	-0.006
	(-1.70)	(-0.40)	(-0.45)	(-0.43)	(0.05)	(-0.67)
$MarketCapShare_{i,t}$	-0.317	-0.668***	-0.659***	-0.666***	0.715***	-0.035
	(-1.09)	(-3.00)	(-2.96)	(-3.00)	(-2.83)	(-0.63)

Table 2.2 (Continued)

	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$
	(1)	(2)	(3)	(4)	(5)	(6)
	TotalSentiment	HRSentiment	LRSentiment	Both Sentiments	Rank at Post	Rank at Post, Random Effect
MarketCapShare$_{i,t}$	-0.317	-0.668***	-0.659***	-0.666***	0.715***	-0.035
	(-1.09)	(-3.00)	(-2.96)	(-3.00)	(-2.83)	(-0.63)
WeekDummy	√	√	√	√	√	√
#Obs.	20,738	9,130	9,130	9,130	6,575	6,575

Note: *** p<0.01, ** p<0.05, * p<0.1.

2.3.2 Incentive Hierarchy and Prediction Accuracy: Evidence from the Bitcoin Markets

As we mentioned before, the discussions in Altcoin-related threads are not organized into different topics, so it is likely that there is a high proportion of noise information. By comparison, Bitcointalk.org categorizes Bitcoin-related discussions into different sections according to the contents, and the discussion section we select for our analysis (Speculation) shall contain more relevant information about Bitcoin pricing. Therefore in this section, we try to verify our first prediction using Bitcoin-related social media discussions.

We collected Bitcoin-related discussions from 3,372 threads from the speculation discussion board from 2015/2/19 to 2017/2/17. Different from the analysis on Altcoins, we constructed the thread-day networks (i.e., separate networks are created for each thread/day) in order to form a panel data set. A total of 57,063 thread-day networks are constructed. Thread-day networks are used due to two considerations: (1) different threads are likely to focus on different topics, and each topic has a unique impact on future price movements, therefore it is reasonable to examine them as cross sections, and (2) financial information is very time-sensitive, the focus and value of even the same topic may vary significantly over different days.

We compare the information posted by users with high rank and users with low rank in terms of prediction accuracy using the following model specification:

$$R_{t+1}=\alpha+\alpha_t+\beta_1 HRSentiment_{it}+\beta_2 LRSentiment_{it}+\delta X+\eta_{it} \quad (2)$$

In Equation (2), i is the thread index. $HRSentiment_{i,t}$ is the aggregate sentiment extracted from social media discussions posted by high rank users for Bitcoin on day t in thread i. $LRSentiment_{i,t}$ is the aggregate sentiment extracted from social media discussions posted by low rank users for Bitcoin on day t in thread i. The time dummy α_t (weekly dummy) controls for the differences in the returns in different time periods. X contains the intraday return R_t, the one-day lagged return R_{t-1}, the two-day lagged return R_{t-2}, and the logarithm of thread-day post count $Ln(PostCount_t)$.

A random effects model is chosen over a fixed effects model because the unobserved disturbance for each thread is more probable to be random rather than fixed in different time periods. Our choice is based on the following two observations. First, participants of the same thread on different days keep changing. Every day, some new authors join the discussions and some old authors leave. This leads to fast-changing dynamics in participating members and their collective wisdom as well. Second, the topic focus of the same thread also changes over time. As new information emerges, discussions also evolve and move from one topic to another. As a result, the unobserved impact of the thread on price movement must also be changing over time. That explains why we cannot represent this unobserved impact with a fixed value.

The estimate results are presented in Table 2.3. The coefficient estimates on $HRSentiment_{i,t}$ are not statistically significant across all model specifications. Therefore, the sentiment extracted from high rank social media users does not predict the future Bitcoin price movement accurately. In contrast, the coefficient estimates on $LRSentiment_{i,t}$ are

statistically significant at least at the 5% level in all five models. Both the evidence from the Bitcoin market and the evidence from the Altcoin market point to superior predictive power from low rank users.

2.3.3 The Implication of Social Media Incentive Hierarchy on the Spillover Effect

In this section, we continue to investigate how incentive hierarchies affect the spillover effects. The finance literature has well documented the phenomenon that new information about a focal firm can affect its intra-industry rivals. Here we try to analyze if the spillover effect exists in the crypto currency market and if so, whether it is mainly caused by high rank users with greater visibility in the online community.

Specifically, we test our hypothesis by studying how information from Bitcoin-related discussion board spills over to the Altcoin markets and how the spillover effect varies for different user groups.

We organize our analysis around the following model specification:

$$R_{i,t+1}=\alpha+\beta_1 Alt\,HRSentiment_{i,t}+\beta_2 Alt\,LRSentiment_{i,t}+\beta_3 Btc\,HRSentiment_t+\beta_2 Btc\,LRSentiment_t \quad (3)$$

As before, the dependent variable $R_{i,t+1}$ is the next-day return for Altcoin *i*, $Alt\,HRSentiment_{i,t}$ is the aggregate sentiment from high rank users for Altcoin *i* on day *t*. $Alt\,LRSentiment_{i,t}$ is the aggregate sentiment from low rank users for Altcoin *i* on day *t*. Similarly, $Btc\,HRSentiment_t$ is the day *t* aggregate sentiment from high rank

Table 2.3 – Predictive Power of Social Media Users with Different Incentive Hierarchy Rank: Bitcoin

	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)	(5)
Badge Used	Final Rank	Final Rank	Final Rank	Rank at Post	Rank at Post
High Badge User Cutoff	Full Member Points > 120	Full Member Points > 120	Hero Member Points > 480	Full Member Points > 120	Full Member Points > 120
PostCount Threshold	> 15	20	> 20	> 15	> 20
$HRSentiment_t$	0.002	-0.004	0.074	0.010	0.016
	(0.02)	(-0.02)	(0.52)	(0.14)	(0.15)
$LRSentiment_t$	-0.083**	-0.142***	-0.203**	-0.088**	-0.145***
	(-1.96)	(-3.62)	(-2.22)	(-2.13)	(-3.76)
$BTCR_t$	-0.296***	-0.265***	-0.294***	-0.295***	-0.268***
	(-9.27)	(-6.09)	(-8.84)	(-9.09)	(-6.14)
$BTCR_{t-1}$	-0.323***	-0.337***	-0.340***	-0.326***	-0.344***
	(-12.62)	(-9.99)	(-12.71)	(-12.51)	(-10.17)
$BTCR_{t-2}$	-0.304***	-0.328***	-0.317***	-0.308***	-0.337***
	(-11.41)	(-9.20)	(-11.34)	(-11.50)	(-9.48)

Table 2.3 (Continued)

	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}	R_{t+1}
	(1)	(2)	(3)	(4)	(5)
Badge Used	Final Rank	Final Rank	Final Rank	Rank at Post	Rank at Post
High Badge User Cutoff	Full Member Points > 120	Full Member Points > 120	Hero Member Points > 480	Full Member Points > 120	Full Member Points > 120
PostCount Threshold	> 15	20	> 20	> 15	> 20
Log(PostCount$_t$)	0.001 (0.20)	-0.001 (-0.27)	-0.001 (-0.28)	0.001 (0.18)	-0.001 (-0.23)
WeekDummy	√	√	√	√	√
#Obs.	1,893	1,181	1,554	1,853	1,169

Note: *** p<0.01, ** p<0.05, * p<0.1.

users in Bitcoin discussion board, and $BtcLRSentiment_t$ is the day t aggregate sentiment from low rank users in Bitcoin discussion board.

We present our results in Table 2.4. In Column (1), we first look at whether the aggregate sentiments from Altcoin and Bitcoin discussions help predict the next day Altcoin return. We can infer that the daily aggregate Altcoin discussion sentiment does not predict the next-day Altcoin return, and more importantly, the sentiment from Bitcoin discussions does not seem to spill over to the Altcoin market.

In Column (2) through Column (4), we break down the sentiments into high rank user sentiment and low rank user sentiment. First, the negative and statistically significant coefficient estimates on $AltLRSentiment_{i,t}$ reassure our results in Section 4.1 and 4.2 that inactive low rank users are the providers of more value-relevant information.

The row 5 and row 6 present our results for the spillover effects. The positive and statistically significant coefficient estimates on $BtcHRSentiment_{i,t}$ point to the superior spillover effect (from the Bitcoin-related discussion board to the Altcoin market) from the high rank users due to their better visibility among peers. While both the positive spillover effect (contagion) and negative spillover effect (competition effect) are documented in the finance literature, here we found the negative spillover effect. In other words, negative sentiments or "bad news" about Bitcoin will make its competitors better off (or higher returns). The coefficient estimates on $BtcHRSentiment_{i,t}$ in Column (4) of Table 2.4 is 1.397, meaning that when there is 1% more negative words in the Bitcoin related discussion, the Altcoin return will be 1.397% higher.

Table 2.4 – Spillover Effects

	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$
	(1)	(2)	(3)	(4)
Badge Used		Final Rank	Final Rank	Rank at Post
High Badge User Cutoff_Altcoin		Full Member, Points > 120	Full Member, Points > 120	Full Member, Points > 120
High Badge User Cutoff_Bitcoin		Full Member, Points > 120	Hero Member, Points > 480	Full Member, Points > 120
$AltTotalSentiment_{i,t}$	0.005			
	(0.13)			
$BtcTotalSentiment_{t}$	0.519			
	(0.57)			
$AltHRSentiment_{i,t}$		0.177	0.175	0.176
		(1.49)	(1.48)	(1.49)
$AltLRSentiment_{i,t}$		-0.154**	-0.152**	-0.153**
		(-2.05)	(-2.03)	(-2.03)
$BtcHRSentiment_{t}$		1.425*	1.785**	1.397**
		(1.71)	(2.51)	(2.15)

Table 2.4 (Continued)

	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$
	(1)	(2)	(3)	(4)
Badge Used		Final Rank	Final Rank	Rank at Post
High Badge User Cutoff_Altcoin		Full Member, Points > 120	Full Member, Points > 120	Full Member, Points > 120
High Badge User Cutoff_Bitcoin		Full Member, Points > 120	Hero Member, Points > 480	Full Member, Points > 120
$BtcLRSentiment_t$		-0.123	-0.194	-0.214
		(-0.48)	(-0.36)	(-0.31)
$ALTR_{i,t}$	-0.037***	-0.028***	-0.027***	-0.027***
	(-5.45)	(-3.16)	(-3.14)	(-3.16)
$ALTR_{i,t-1}$	-0.028***	-0.018***	-0.018***	-0.018***
	(-4.03)	(-3.99)	(-3.99)	(-4.01)
$ALTR_{i,t-2}$	-0.013*	-0.014***	-0.013***	-0.013***
	(-1.84)	(-2.89)	(-2.86)	(-2.86)
$BTCR_{i,t}$	-0.302***	-0.312***	-0.309***	-0.309***
	(-5.38)	(-5.84)	(-5.78)	(-5.78)

Table 2.4 (Continued)

	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$
	(1)	(2)	(3)	(4)
Badge Used		Final Rank	Final Rank	Rank at Post
High Badge User Cutoff_Altcoin		Full Member, Points > 120	Full Member, Points > 120	Full Member, Points > 120
High Badge User Cutoff_Bitcoin		Full Member, Points > 120	Hero Member, Points > 480	Full Member, Points > 120
$BTCR_{i,t-1}$	-.166***	-0.187***	-0.185***	-0.186***
	(-2.93)	(-3.52)	(-3.46)	(-3.49)
$BTCR_{i,t-2}$	-0.140**	-0.166***	-0.165***	-0.164***
	(-2.46)	(-3.12)	(-3.10)	(-3.10)
$Log(PostCount_{i,t})$	0.012*	0.0004	0.0006	0.0006
	(1.82)	(0.08)	(0.11)	(0.09)
$Log(AuthorCount_{i,t})$	-0.014*	-0.003	-0.004	-0.004
	(-1.70)	(-0.43)	(-0.47)	(-0.45)
$MarketCapShare_{i,t}$	-0.317	-0.667***	-0.666***	-0.667***
	(-1.09)	(-3.00)	(-3.00)	(-3.00)
WeekDummy	√	√	√	√

Table 2.4 (Continued)

	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$	$R_{i,t+1}$
	(1)	(2)	(3)	(4)
Badge Used		Final Rank	Final Rank	Rank at Post
High Badge User Cutoff_Altcoin		Full Member, Points > 120	Full Member, Points > 120	Full Member, Points > 120
High Badge User Cutoff_Bitcoin		Full Member, Points > 120	Hero Member, Points > 480	Full Member, Points > 120
#Obs.	20,738	9,130	9,130	9,130

2.4 Conclusion

In this chapter, we analyze the implication of social media incentive hierarchies on social media users' posting motivation and empirically test the influence on the predictive accuracy. For active users with high ranks, because of their reduced motivation after obtaining the high ranks and frequent use of social media for the purpose of socialization, their posts contain a high proportion of noise information. However, high badges users do enjoy better visibility. This research helps understand the drawbacks of activity-based social media incentive hierarchy systems.

REFERENCES

Akhigbe, Aigbe, and Anna D. Martin. 2000. "Information-signaling and Competitive Effects of Foreign Acquisitions in the US," *Journal of Banking & Finance*(24:8), pp. 1307-1321.

Akhigbe, Aigbe, Jeff Madura, and Anna D. Martin. 2015. "Intra-industry Effects of Negative Stock Price Surprises," *Review of Quantitative Finance and Accounting*(45:3), pp. 541-559.

Anderson, Ashton, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2013. "Steering User Behavior with Badges," In *Proceedings of the 22nd international conference on World Wide Web*, pp. 95-106.

Antweiler, Werner, and Frank, Murray Z. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards," *The Journal of Finance* (59:3), pp. 1259-1294.

Antin, Judd, and Elizabeth F. Churchill. 2011. "Badges in Social Media: A Social Psychological Perspective," CHI 2011 Gamification Workshop Proceedings. New York, NY: ACM.

Bramoullé, Yann, and Kranton, Rachel. 2007. "Public Goods in Networks," *Journal of Economic Theory* (135:1), pp. 478-494.

Blei, David M, Ng, Andrew Y, and Jordan, Michael I. 2003. "Latent Dirichlet Allocation," *Journal of Machine Learning Research* (3), pp. 993-1022.

Chen, Hailiang, De, Prabuddha, Hu, Yu Jeffrey, and Hwang, Byoung-Hyoun. 2014. "Wisdom of Crowds: The Value of Stock Opinions Transmitted through Social Media," *Review of Financial Studies* (27:5), pp. 1367-1403.

Chen, Sheng-Syan, Kim Wai Ho, and Kueh Hwa Ik. 2005. "The Wealth Effect of New Product Introductions on Industry Rivals," *The Journal of Business*(78:3), pp. 969-996.

Cheong, C., Cheong, F. and Filippou, J. 2013."Quick Quiz: A Gamified Approach for Enhancing Learning,". In *PACIS*, pp. 206.

Coleman, J. 1990. Foundations of Social Theory, Cambridge, Mass.: *Belknap Press of Harvard University Press*.

Colla, Paolo, and Mele, Antonio. 2010. "Information Linkages and Correlated Trading," *Review of Financial Studies* (23:1), pp. 203-246.

Das, Sanjiv R, and Chen, Mike Y. 2007. "Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science* (53:9), pp. 1375-1388.

Davis, Angela K, Piger, Jeremy M, and Sedor, Lisa M. 2012. "Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language," *Contemporary Accounting Research* (29:3), pp. 845-868.

Denny, Paul. 2013. "The Effect of Virtual Achievements on Student Engagement." In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 763-772.

Deterding S, Dixon D, Khaled R, Nacke L. 2011. "From Game Design Elements to Gamefulness: Defining Gamification," In Proceedings of the 15th international academic MindTrek conference: *Envisioning future media environments*, pp. 9-15.

Dewally, Michael. 2003. "Internet Investment Advice: Investing with a Rock of Salt," Financial Analysts Journal (59:4), pp. 65-77.

DomíNguez, A., Saenz-De-Navarrete, J., De-Marcos, L., FernáNdez-Sanz, L., PagéS, C. and MartíNez-HerráIz, J.J. 2013. "Gamifying Learning Experiences: Practical Implications and Outcomes," *Computers & Education*(63) , pp. 380-392.

Elliott, R. Stephen, Michael J. Highfield, and Mark Schaub. 2006. "Contagion or Competition: Going Concern Audit Opinions for Real Estate firms," *The Journal of Real Estate Finance and Economics* (32:4), pp. 435-448.

Farzan, Rosta, and Peter Brusilovsky. 2011. "Encouraging User Participation in a Course Recommender System: An Impact on User Behavior," *Computers in Human Behavior*(27:1), pp. 276-284.

Ferris, Stephen P., Narayanan Jayaraman, and Anil K. Makhija. 1997. "The Response of Competitors to Announcements of Bankruptcy: An Empirical Examination of

Contagion and Competitive Effects," *Journal of corporate finance*(3:4), pp. 367-395.

Goes, Paulo B., Chenhui Guo, and Mingfeng Lin. 2016. "Do Incentive Hierarchies Induce User Effort? Evidence from an Online Knowledge Exchange," *Information Systems Research* (27:3), pp. 497-516.

Goins, Sheila, and Thomas S. Gruca. 2008. "Understanding Competitive and Contagion Effects of Layoff Announcements," *Corporate Reputation Review*(11:1), pp. 12-34.

Grant, Scott, and Buddy Betts. 2013. "Encouraging User Behaviour with Achievements: an Empirical Study." In *Mining Software Repositories*, pp. 65-68.

Gray, Wesley R, and Kern, Andrew E. 2011. "Talking Your Book: Social Networks and Price Discovery," *Available at SSRN 1767452).*

Grossman, Sanford J, and Stiglitz, Joseph E. 1980. "On the Impossibility of Informationally Efficient Markets," *The American economic review* (70:3), pp. 393-408.

Han, Bing, and Yang, Liyan. 2013. "Social Networks, Information Acquisition, and Asset Prices," *Management Science* (59:6), pp. 1444-1457.

Hausman, J. A. 1978. "Specification Tests in Econometrics," *Econometrica* (46:6), pp. 1251-1271.

Helwege, Jean, and Gaiyan Zhang. 2015. "Financial Firm Bankruptcy and Contagion," *Review of Finance*(20:4), pp. 1321-1362.

Hsu, Hung-Chia, Adam V. Reed, and Jorg Rocholl. 2010. "The New Game in Town: Competitive Effects of IPOs," *The Journal of Finance*(65:2), pp. 495-528.

Jung, J. H., Christoph Schneider, and Joseph Valacich. 2010. "Enhancing the Motivational Affordance of Information Systems: The Effects of Real-time Performance Feedback and Goal Setting in Group Collaboration Environments," *Management Science*(56:4), pp. 724-742.

Kollock, Peter, and Marc Smith. 1996. "Managing the Virtual Commons," *Computer-mediated communication: Linguistic, Social, and Cross-cultural Perspectives*, pp. 109-128.

Lang, Larry HP, and RenéM Stulz. "Contagion and Competitive Intra-industry Effects of Bankruptcy Announcements: An Empirical Analysis," *Journal of financial economics*(32:1), pp. 45-60.

Laux, Paul, Laura T. Starks, and Pyung Sig Yoon. 1998. "The Relative Importance of Competition and Contagion in Intra-industry Information Transfers: An Investigation of Dividend Announcements," *Financial Management*, pp. 5-16.

Loughran, Tim, and McDonald, Bill. 2011. "When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks," *The Journal of Finance* (66:1), pp. 35-65.

Otchere, Isaac. 2007. "Does the Response of Competitors to Privatization Announcements Reflect Competitive or Industry-wide Information Effects? International Evidence," *Journal of Empirical Finance*(14:4), pp. 523-545.

Ozsoylev, Han N, and Walden, Johan. 2011. "Asset Pricing in Large Information Networks," *Journal of Economic Theory* (146:6), pp. 2252-2280.

Qiu, Liangfei, Cheng, H Kenneth, and Pu, Jingchuan. 2016. "'Hidden Profiles' in Corporate Prediction Markets: The Impact of Public Information Precision and Social Interactions," *MIS Quarterly* forthcoming.

Slovin, Myron B., Marie E. Sushka, and John A. Polonchek. 1999. "An Analysis of Contagion and Competitive Effects at Commercial Banks," *Journal of Financial Economics*(54:2), pp. 197-225.

Solomon, David H. 2012. "Selective Publicity and Stock Prices," *The Journal of Finance* (67:2), pp. 599-638.

Stasser, Garold, and Titus, William. 1985. "Pooling of Unshared Information in Group Decision Making: Biased Information Sampling During Discussion," *Journal of personality and social psychology* (48:6), p. 1467.

Tawatnuntachai, Oranee, and Ranjan D'Mello. 2002. "Intra-industry Reactions to Stock Split Announcements," *Journal of Financial Research*(25:1), pp. 39-57.

Tetlock, Paul C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market," *The Journal of Finance* (62:3), pp. 1139-1168.

Tetlock, Paul C, Saar-Tsechansky, Maytal, and Macskassy, Sofus. 2008. "More Than Words: Quantifying Language to Measure Firms' Fundamentals," *The Journal of Finance* (63:3), pp. 1437-1467.

Thom, Jennifer, David Millen, and Joan DiMicco. 2012. "Removing Gamification from an Enterprise SNS," In *Proceedings of the acm 2012 conference on computer supported cooperative work*, pp. 1067-1070.

Tumarkin, Robert, and Whitelaw, Robert F. 2001. "News or Noise? Internet Postings and Stock Prices," *Financial Analysts Journal* (57:3), pp. 41-51.

Wasko, M. McLure, and Samer Faraj. 2000. "“It Is What One Does”: Why People Participate and Help Others in Electronic Communities of Practice," *The Journal of Strategic Information Systems* (9:2), pp. 155-173.

Wasko, Molly McLure, and Faraj, Samer. 2005. "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *MIS quarterly* (29:1), pp. 35-57.

Witt, Maximilian, Christian Scheiner, and Susanne Robra-Bissantz. 2011. "Gamification of Online Idea Competitions: Insights from an Explorative Case," *Informatik Schafft Communities*.

Yoo, Eunae, Rand, William, Eftekhar, Mahyar, and Rabinovich, Elliot. 2016. "Evaluating Information Diffusion Speed and Its Determinants in Social Media Networks During Humanitarian Crises," *Journal of Operations Management* (45), pp. 123-133.